Process Support for Learning Tasks in Multimedia Practicals.

Process Support for Learning Tasks in Multimedia Practicals

Process Support for Learning Tasks in Multimedia Practicals

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de Open Universiteit Nederland op gezag van de rector magnificus prof. dr. ir. F. Mulder ten overstaan van een door het College voor promoties ingestelde commissie in het openbaar te verdedigen

op vrijdag 5 maart 2004 te Heerlen om 16:00 uur precies

door
Robertus Joseph Nadolski
geboren op 4 maart 1961 te Sittard

Promotores:
Prof. dr. J. J. G. van Merriënboer, Open Universiteit Nederland
Prof. dr. P. A. Kirschner, Open Universiteit Nederland

Overige leden van de beoordelingscommissie:
Prof. dr. dr. h.c. mult. F. Achtenhagen, Georg-August-Universität Göttingen
Prof. dr. H. P. A. Boshuizen, Open Universiteit Nederland
Prof. dr. P. R. J. Simons, Universiteit Utrecht
Dr. mr. J. Wöretshofer, Open Universiteit Nederland

www.ou.nl/otecresearch
ISBN 90-9017710-8

Omslag: Jeroen Berkhout, Open Universiteit Nederland m.m.v. Sem Wigman

Druk: Datawyse Maastricht

"The journey of a thousand miles begins with one step" (Lao Tzu)

CONTENTS

Chapter 1 - General introduction

We struggle with rapid developments in our jobs that are becoming increasingly complex and where the solutions to the problems that we encounter are not readily available. Society continues to make stronger demands on flexible problem solving behavior based upon applying complex cognitive skills. Acquiring these complex cognitive skills can only be accomplished through a *complex learning* process where knowledge, skills, and attitudes are acquired and integrated and where these are coordinated during task execution. Only then can we acquire those complex skills that aim at transfer of what is learned in school to daily life or work settings. The challenging question for education is: How can we help students acquire these complex cognitive skills?

Modern instructional theories focus increasingly on authentic learning tasks based on real-life tasks as the paramount condition for learning (Merrill, 2002; Reigeluth, 1999; van Merriënboer & Kirschner, 2001). A considerable risk with using such authentic tasks is that they are often too difficult for novice learners to deal with as a whole. A common solution for this problem is to provide *process support* by (a) splitting the problem solving process of whole learning tasks into smaller phases and presenting them to the learners, and (b) offering driving questions to help learners carry out those phases. The amount of process support must, in turn, be optimized for efficient and effective learning.

Using authentic learning tasks is a challenging experience for instructional designers, especially in distance education. Where traditional universities can use internships, laboratories, and field trips, distance universities such as the Open University of the Netherlands lack such facilities and are forced to look for suitable replacements such as simulations or Multimedia Practicals. Although instructional designers at the Open University of the Netherlands have a lot of experience with such Multimedia Practicals and use various methods to provide support within them (Gerrichhauzen et al., 1998; Hoogveld, et al., 1997; Hommes et al., 2000; Huysse et al., 1998; Ivens et al., 1998; Leinders et al., 1993; Wöretshofer et al., 2000), systematic research on the effects of such methods is lacking. This thesis addresses the effects of process support (via the number of phases and the provision of driving questions) on task performance and task efficiency within Multimedia Practicals.

This general introduction discusses the frame of reference underlying the use of process support to facilitate acquiring complex cognitive skills. First, the process of acquiring complex cognitive skills is discussed. This is followed by a discussion of the design requirements for acquiring such skills through Multimedia Practicals. Finally, the design of process support in Multimedia Practicals is treated in greater detail and research questions on the provision of process support are presented. The chapter concludes with an overview of the content of this thesis.

Acquiring complex cognitive skills

Complex cognitive skills are skills for which the learner must invest considerable time and effort to acquire an acceptable mastery level and for which qualitative differences in performance exist between novices and experts. Exemplary complex skills are: diagnosing a particular disease, selecting a suitable job applicant, modeling stress-factors that cause mental overload in workers, or preparing a plea to be held in court. The essence of a complex cognitive skill is that its mastery involves coordination and integration of its constituent skills and not simply the mastery of those separate constituent skills. A person who has mastered such a skill can apply it in a variety of realistic situations. Such application involves both flexible problem solving and smoothly carrying out the necessary skills.

For flexible problem solving behavior, an expert often relies on *domain-based cognitive strategies* that are represented within problem schemas in long-term memory (see, e.g., Clark, 1998; Gagné, Yekovich, & Yekovich, 1993). In a sense, an experts' approach to problem solving is a matter of *recognizing* patterns previously experienced and *matching* these patterns to corresponding aspects of the problem at hand. In fact, because of their extensive elaborated subject matter knowledge in a domain, experts are able to use robust domain-based cognitive strategies in their problem solving that they have derived themselves. Novices, on the other hand, do not possess sufficiently elaborated subject matter knowledge to permit such derivations (e.g., Chase & Simon, 1973a, 1973b; Chi, Glaser, & Rees, 1982; Larken et al., 1980), and are consequently forced to apply more general cognitive strategies that lack both efficiency and power in solving domain-based problems. In addition, early research on human problem solving (e.g., Newell & Simon, 1972) has made clear that novices performing complex tasks utilize cognitive strategies that keep the information processing demands of the situation within the bounds of their limited working memory capacity. This, however, often leads to nonefficient learning (Craik & Lockhart, 1972; Sternberg & Frensch, 1991).

Good instruction should not simply offer the same - although successful - time-consuming and effortful road to expertise that has been taken by experts, but should offer an alternative road to foster the development of domain-based cognitive strategies by novices. Well-designed whole-task approaches to instruction may provide this alternative.

Design requirements for Multimedia Practicals for acquiring complex cognitive skills

Most cognitive skills, especially more complex ones, consist of a number of simpler constituent cognitive skills. Traditional Instructional Design models (e.g., Dick & Carey, 1979; Romiszowski, 1981) are not suitable for designing learning environments for acquiring such complex cognitive skills because they focus too strongly on the acquisition and the training of constituent skills instead of the *complete* complex cognitive skill. Instructional Design models using whole-task approaches combined with Multimedia Practicals could overcome this shortcoming.

Whole-task approaches

The emphasis on authentic, whole tasks can be found in both practical educational approaches such as problem-based learning, project-centered education, and competency-based learning and theoretical models such as Collins, Brown, and Newman's (1989) theory of cognitive apprenticeship learning, Nelson's (1999) theory of collaborative problem solving, and Jonassen's (1999) theory of constructive learning environments. Whole-task approaches (e.g., van Merriënboer, 1997) emphasize the coordination and integration of constituent skills from the very beginning, and stress that learners should quickly develop a holistic vision of the whole task that is gradually embellished during the training. In early stages of whole-task approaches much support is provided. This support is then faded as the learner becomes more proficient. A considerable risk with using authentic, whole tasks is that they are often too difficult for learners to deal with. To alleviate this problem, Achtenhagen (2001) suggests pedagogically modeling an already conceptually modeled reality. Pedagogically modeling the model (i.e., didactic specification, Resnick, 1976) is often achieved through the use of two process support mechanisms, namely (a) segmenting the whole learning task into smaller task assignments and thereby dividing the problem solving process into *phases* which are presented via a process worksheet, and (b) providing *driving questions* to help learners in carrying out the phases (Land, 2000). In other words, after modeling reality, pedagogical modeling streamlines the problem-solving process of the whole learning task by structuring it into phases and scaffolding the phases via driving questions. Process support is pivotal in fostering the acquisition of domain-based cognitive strategies and can be expected to promote transfer of learning.

Multimedia Practicals

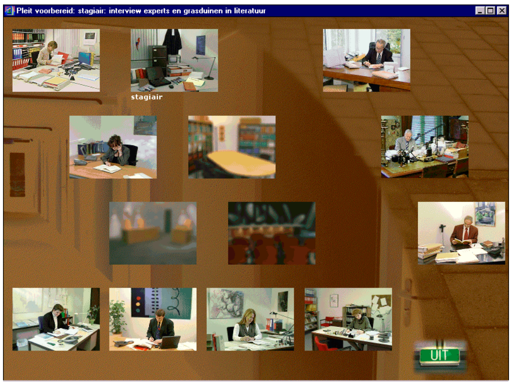
Multimedia Practicals developed at the Open University of the Netherlands are *simulated* task environments, modeled after realistic situations that offer the aforementioned sequence of whole learning tasks. Situated learning (Brown, Collins, & Duguid, 1989, Westera & Sloep, 1998) emphasizes that learning environments need to offer realistic situations where learning through meaningful practice takes place; the premise being that acquisition of complex skills is context-dependent and occurs most effectively in a relevant context (Anderson, 1982, 1993; Brown et al., 1989; Kirschner, van Vilsteren, Hummel, & Wigman, 1997; Kolb, 1984; van Parreren, 1987). Multimedia Practicals attempt to provide realistic situations where meaningful practice takes place in an electronic *self-contained* learning environment (i.e., all necessary support is embedded in the environment). The relevant context is often provided through a virtual working environment modeled after the environment where such tasks normally are conducted. The problems typically have a well-defined begin state, many possible pathways to reach a solution, and (usually) not very well-defined end states (are ill-structured), and often require 10 to 20 hours to complete.

This thesis concentrates on studying the effects of *process support* (i.e., support for acquiring domain-based cognitive strategies) in Multimedia Practicals within a whole-task instructional sequence. The research context was an adaptation of the Multimedia Practical *Preparing a plea* (Wöretshofer et al., 2000) within the Law Faculty.

***Research context: Multimedia Practical* Preparing a plea**

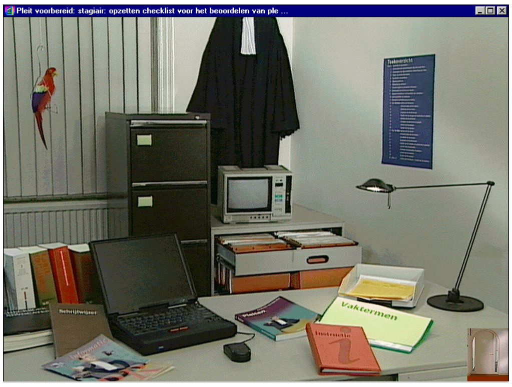
In the Multimedia Practical *Preparing a plea* the learner is a trainee in a virtual law firm. The trainee first studies a general introduction to pleading a case, in which supportive information and various support tools are provided. Supportive information is helpful to the learning and execution of problem-solving aspects of learning tasks. It is often regarded as 'the theory' by teachers. The support tools include examples of lawyers conducting a plea, discussions of ethical issues in pleading a case, numerous tips on the communicative aspects in pleading a case, and judicial-procedural aspects of plea preparation. During this general introduction, the trainee receives several assignments to guide the study of the theory as well as support from a senior (virtual) employee of this firm, the coach (see Figure 1.1a). The trainee can make use of standard office equipment and can visit other places in the firm, such as experts' offices (see Figure 1.1b). The trainee can, for example, study the legal backgrounds of different cases in a file cabinet, observe and analyze other pleas using a "plea checker" (see Figure 1.1c), make electronic notes, attend staff meetings, and consult experts. After this general introduction, the trainee must prepare pleas for various cases. The case files are available within a (virtual) office (see Figure 1.1d). The coach provides task assignments with feedback for each segment in the whole task. Finally, the trainee conducts the prepared pleas outside the Multimedia Practical, an actual simulated courtroom.

1.1a



1.1b

1.1c



1.1d

Figure 1.1
Various screen dumps taken from the Multimedia Practical *Preparing a plea.* Explanations are provided in the text.

Designing process support in Multimedia Practicals

As stated, Multimedia Practicals make use of two process support mechanisms namely process worksheets that divide the process of carrying out the whole task into phases and driving questions that help the carrying out of a phase. Both kinds of process support are domain-specific and tuned to the task at hand and are very different from content-free heuristics (i.e., general problem solving methods) or content-free questions (e.g., in case of writing a report: asking if the person checked the spelling, or added an index).

Research questions

The central research questions with respect to the provision of process support are:
(1) does the number of phases influence performance on the task and the efficiency of carrying out the task, and if so, in what way?
(2) do driving questions positively influence performance on the task and the efficiency of carrying out the task?

Number of phases

Optimizing the number of phases and providing an accompanying process worksheet brings proficient execution of the learning tasks within the reach of the learners' capabilities. *Process worksheets* (van Merriënboer, 1997) offer a way to provide the phases to the learners and guide them through the problem solving process of the whole learning task; they provide a Systematic Approach to Problem Solving for the whole learning task. As an example (see Figure 1.2), the whole task of "preparing & pleading a case in court" has been segmented into seven meaningful task assignments, each of which is represented by a phase in the process worksheet. Complex tasks divided into too few phases are often too difficult and mentally demanding for learners to carry out, which hampers learning and subsequent transfer. Learners may either not accurately process the necessary information because they are overwhelmed by the difficulty of the task (i.e., cognitive overload) or they may revert to surface processing (i.e., superficial, non-meaningful learning) in order to keep their cognitive load within the threshold limit (Craik & Lockhart, 1972; Sternberg & Frensch, 1991).

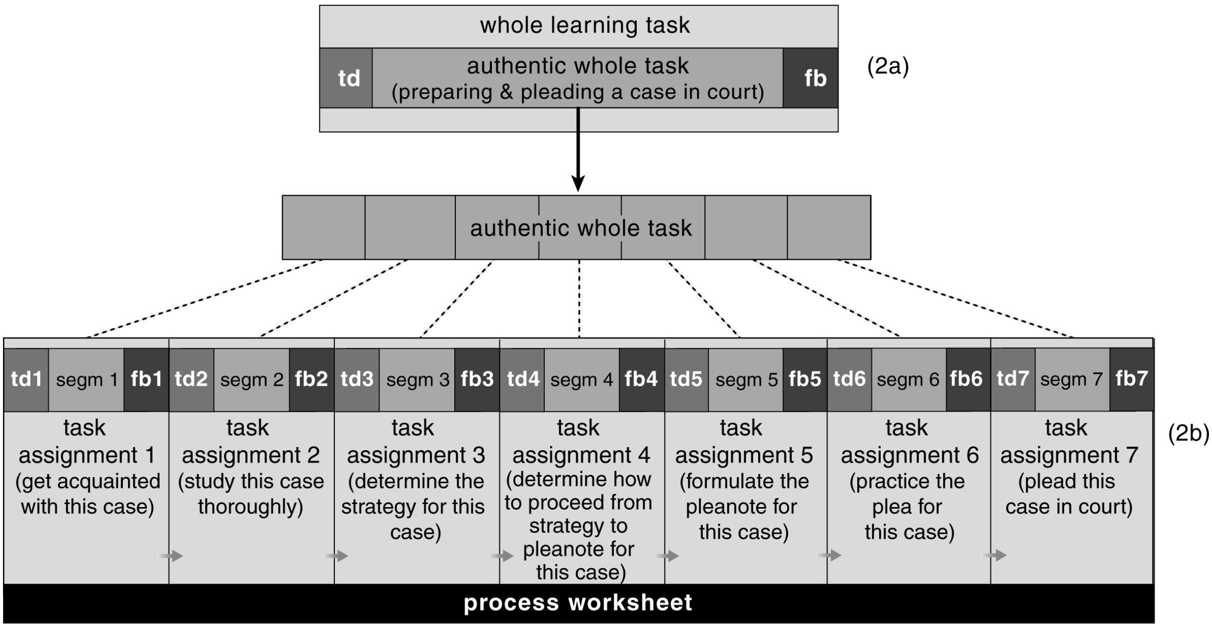


Figure 1.2
Authentic whole task, segmented in task assignments.
(2a) A non-segmented whole learning task consists of an authentic whole task preceded by task description (td) and followed by feedback (fb). (2b) Each task assignment (numbered 1 through 7) also consists of a task description, an authentic subtask (i.e., segment from the authentic whole task) and feedback. Each task assignment represents a phase in the problem solving process of the whole task. The process worksheet presents the phases in the problem solving process of the whole task. Task descriptions can include driving questions for carrying out the activities within a phase.

Tasks divided into too many phases may also hamper learning because of the non-coherency caused by redundant information between phases and/or an excess of details making them mentally demanding (Mayer & Moreno, 2002). In addition, learners may regard the phases as being too specific to the learning task in question, preventing them from constructing the generalizations or abstract schemas necessary for transfer of learning to occur. Thus, like many other instructional design problems (see Clark, 1999), determining the number of phases is an instructional design problem that requires a solution through optimization that, in turn, needs determination of objective task complexity.

Driving questions

Driving questions scaffold the problem solving process within the phases of whole learning tasks. They are meant to help the learners in carrying out the activities in a phase, for instance, by referring them to information resources in a correct and efficient way, by activating relevant prior knowledge, or by suggesting relevant procedures and principles. In this thesis, driving questions guide learners in using appropriate resources and selecting relevant elements from them to apply in the problem solving process for one particular phase (see Table 1.1).

Table 1.1
Driving questions for 'Get acquainted with a law case' (i.e., task description for segment 1 in Figure 1.2)

What field of law does this case belong to?
Who are the parties in this case?
Whose representative will you be?
Who is bringing the case before court?
What kind of procedure should be used?
What kind of judge will try this case?
What about the territorial jurisdiction of the judge?
What is this case roughly about?
Is there a previous judicial/legal history, and if so, is it relevant for this case?
What do you already know about the subject matter in this case?
Is the file complete?
What legal aspects could be of importance in this case?
What clues does the case offer at first sight?
What is the tentative goal of your plea?

Note that these questions are not specific to a certain case, but are specific to a certain domain

Driving questions are given in an arbitrary order at the start of a phase, are kept available during task execution, and are meant to facilitate the problem solving process for the current phase. They are expected to enhance performance because they focus on essential problem solving skills (see, e.g., Smith & Ragan, 1999).

Content overview of the thesis

Chapter 2 presents an Instructional Design model for Multimedia Practicals for acquiring complex cognitive skills. This model was used for the design of the adapted versions of the Multimedia Practical, which were included as experimental conditions in the studies on the provision of process support.

Chapter 3 describes a study carried out to develop a reliable, valid and easy-to-use measurement instrument for objectively rating learning task complexity. This task complexity measurement instrument is used both in the Instructional Design model for Multimedia Practicals as well as for determining the complexity of the tasks in the studies on process support.

Chapter 4 reports on a study of how the number of phases influences task performance and task efficiency. Task performance relates to the physically not perceptible results of learning (see, e.g. Boekaerts & Simons, 2003). Task efficiency relates to the investment made to reach those results in terms of task motivation (Bonner, 1994; Maynard & Hakel, 1997), mental effort (Paas & van Merriënboer, 1994), and time on task (Karweit, 1984).

Chapter 5 reports on a study of how driving questions affect task performance and task efficiency, and whether there was an interaction between the number of phases and the availability of driving questions.

Chapter 6, the final chapter of this thesis, presents a general discussion of the design approach and discusses the results of the empirical studies, presents practical implications and guidelines of the research, and gives suggestions for future research.

Chapter 2 – Instructional Design model for Multimedia Practicals*

Learners are often overwhelmed by the complexity of realistic learning tasks, but reducing this complexity through traditional Instructional Design (ID) methods jeopardizes the authenticity of the learning experience. To solve this apparent paradox, a two-phase six-step ID model is presented. Phase 1 consists of cognitive task analysis, where a systematic approach to problem solving (SAP) is identified in conjunction with skill decomposition and determination of task complexity. In the subsequent design phase, inductive micro-level sequencing based on van Merriënboer's four-component ID model is applied where worked-out examples and problems accompanied by process worksheets assure the necessary variability of practice. The number of phases in a multiple-segment whole-task approach - needed for the process worksheets - is determined on the basis of estimated task complexity. A developmental study of the model is illustrated with examples from the domain of Law.

Introduction

In designing competency-based learning environments, the challenge is to facilitate learning while providing authentic tasks. Authentic undecomposed tasks are often too complex for learners to deal with. In this chapter we present an instructional design (ID) model that focuses on optimizing the number of phases in process worksheets in whole-task approaches for acquiring complex, mainly non-recurrent, cognitive skills. The model consists of six steps, namely: skill decomposition, determination of task complexity, identification of systematic approaches to problem solving (SAPs), micro-level sequencing of problems, choosing problem formats, and choosing the number of phases in SAPs to be presented to learners through process worksheets. All are important to facilitate learner task performance in competency-based learning environments. The model is, otherwise stated, concerned with task analysis and design of learning tasks for such environments.

Multimedia Practicals (MmP) provide realistic situations in which meaningful learning through contextualized practice takes place (Brown, Collins, & Duguid, 1989; Westera & Sloep, 1998) and can be regarded as competency-based learning environments. These practicals usually deal with complex skills consisting of an integrated set of constituent skills. Although some constituent skills may be recurrent from problem situation to problem situation (i.e., procedural), non-recurrent constituent skills where the desired behavior is contextually dependent and where transfer to new problem situations should occur are most important. The total skill-set for MmPs is also referred to as the *goal competency*.

Problems within MmPs typically have a well-defined begin state, many possible pathways, and usually not very well-defined end states (are ill-structured). Such problems can be extremely large. An example of such a problem is preparation to plead a case in court (Wöretshofer et al., 2000). Because learners are unfamiliar with the problems posed and thus do not know how to approach them (they do not possess the necessary SAP), the problem-task

---

* Based on: Nadolski, R. J., Kirschner, P. A., van Merriënboer, J. J. G, & Hummel, H. G. K. (2001). A model for optimizing step size of learning tasks in Competency-based Multimedia Practicals. *Educational Technology Research and Development, 49,* 87-103.

representing the goal competency is too complex to achieve in one try. In other words, the "size" of the task is too large. In contrast, provided that the necessary support is given to the learners, the problem-task itself is not so difficult that it cannot be practiced as a whole. Learners have most, if not all, of the essential prior knowledge and skills, but have never combined them in the prescribed manner (the SAP).

The model prescribes six steps in two phases that provide the necessary support to learners. The first three steps form the *cognitive task analysis phase*; the final three steps constitute the *design phase*. Before describing this model, we first elaborate on learning in such practicals.

Learning in Multimedia Practicals

MmPs are typically *simulated* task environments, modeled after realistic situations. The whole learning tasks that learners have to deal with involve the acquisition of complex cognitive skills and are derived from authentic whole tasks. Situational learning (Brown et al., 1989) emphasizes that such environments need to offer realistic situations where learning through meaningful practice takes place; complex skills-learning occurs most effectively in a relevant context. This knowledge construction is context-dependent and cannot be isolated from situations in which it is learned (Anderson, 1982, 1993; Brown et al., 1989; Kirschner, van Vilsteren, Hummel, & Wigman, 1997; Kolb, 1984; van Parreren, 1987). It is assumed that complex learning requires the mindful abstraction of cognitive schemas from concrete experiences. However, the full complexity of real-life tasks typically interferes with such effort-demanding inductive processing. A common solution to preclude this problem is first to conceptually model reality (i.e., simplify it) and second to pedagogically model this model (i.e., make it learnable; Achtenhagen, 2001). Pedagogical modeling (i.e., didactic specification, see Resnick, 1976) can be achieved through *segmentation* of the whole learning task into smaller task assignments and thereby dividing the problem solving process into *phases* which are presented via a process worksheet. In other words, after modeling reality, pedagogical modeling streamlines the problem solving process of the whole learning task as it segments this process into phases (see Figure 2.1). The example presented in Figure 2.1 is taken from the MmP *Preparing a plea* (Wöretshofer et al., 2000).

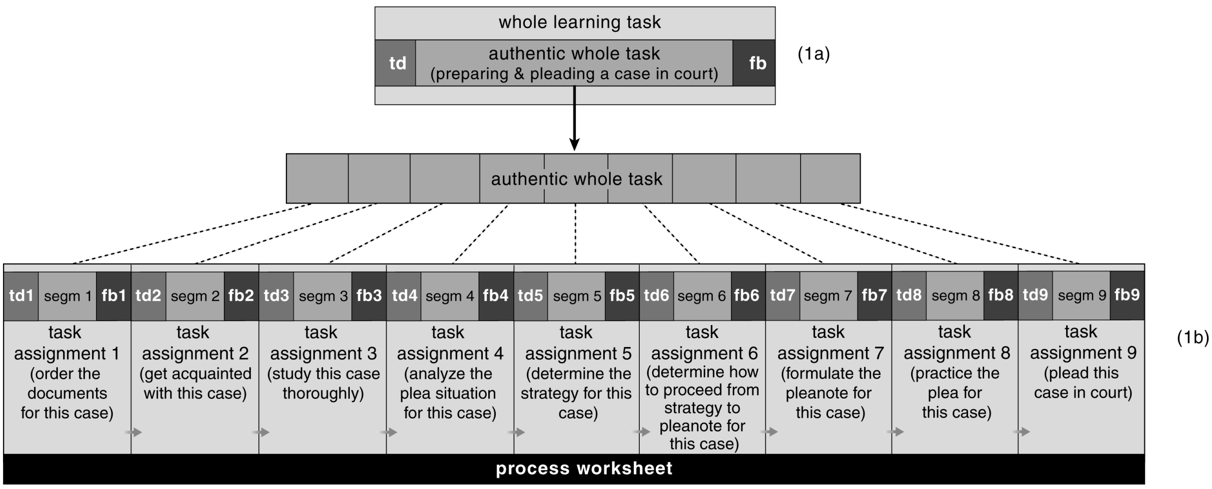


Figure 2.1
Authentic whole task, segmented in task assignments.
(1a) A non-segmented whole learning task consists of an authentic whole task preceded by task description (td) and followed up with feedback (fb). (1b) Each task assignment (numbered 1 through 9) also consists of a task description, an authentic subtask (i.e., segment from the authentic whole task) and feedback. Each task assignment represents a phase in the problem solving process of the whole task. The process worksheet presents the phases in the problem solving process of the whole task.

In this MmP that is available on CD-ROM, the learner is a trainee in a law firm, and must prepare pleas for various cases. The case files are available within an (electronic) office. As trainee, the learner receives support from a senior (electronic) employee of this firm, the coach. This coach introduces how to prepare a plea and comments on various activities (i.e., task assignments of the whole task 'to conduct a plea') that the trainee performs during the preparation phase. The trainee can make use of office equipment and can visit other places in the firm. The trainee can - for example - observe and analyze other pleas using a "plea checker", study legal backgrounds of different pleas, consult experts, and attend staff meetings. Finally, the trainee conducts the prepared pleas in real-life two-day role playing exercises.

Learning in MmPs typically involves acquiring a set of highly interrelated, non-recurrent constituent skills (*goal competencies*) involving a high degree of transfer. Skill performance is based on schema-based behaviors after the training. Problem solving consists of first finding an appropriate problem schema in long-term-memory and then filling this schema with the specific parameters of the problem at hand (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Farr, 1988). The problem schema that is retrieved in a particular case is a crucial determinant of how the problem is solved since it determines both the conceptual knowledge used to elaborate the problem statement and the approaches used to solve the problem (Gagné, Yekovich, & Yekovich, 1993). Complex cognitive skills involve both problem solving and skilled performance; the recurrent constituent skills are driven by automated schemas held in long-term memory.

MmPs anticipate novice problem-solving behavior by offering a process worksheet to guide learners through the problem-solving process instead of over-challenging them to induce their own SAP. SAPs are domain-specific problem-solving strategies with their associated heuristics. Learners in MmPs start as novices and act as apprentices since they have not encountered such problems before (Bedard & Chi, 1992). They, therefore, use certain *novice* strategies when solving those problems. Early research on human problem solving (e.g., Newell & Simon, 1972) has made clear that individuals performing complex tasks utilize heuristics that keep the information processing demands of the situation within the bounds of their limited working memory capacity. This, however, often leads to nonefficient learning, resulting in surface processing instead of deep, meaningful processing (Craik & Lockhart, 1972, Sternberg & Frensch, 1991).

Tasks in MmPs

Tasks in a MmP are performed using a SAP encompassing a sequence of phases with associated subgoals. Each phase is accompanied by a set of heuristics that may be used to reach the subgoals and thus to achieve the goal competency. SAPs represent the needed *strategic knowledge*. In our example, the goal competency is "pleading a case in court". The SAP is provided in Table 2.1. Note that this SAP is nonhierarchic and that the result of each phase is input for the next; iterations between phases are possible. Further note that this SAP specifies the content of the process worksheet depicted in Figure 2.1.

Transfer for pleading a case in court entails the learner's ability to plead a case in any domain of law (e.g., criminal law or civil law) and in any court (e.g., single judge or three judges).

Table 2.1
A systematic approach to problem solving (SAP) for "preparing & pleading case X in court"

Subgoals (phase)	*Heuristics*
1.Order documents for case X	You might try to order the documents chronologically, categorically (e.g., legal documents, letters, notes), or by relevance.
2.Get acquainted with case X	You might answer questions such as "Which subdomain of law is relevant for this case?" or "How do I estimate the chances for my client?"
3.Study case X thoroughly	You might provide answers to questions such as "What is the specific legal question in this case?", "What sections of the law are relevant in this case?" or "What legal consequence is most convenient for my client?"
4.Analyze the situation for a plea for case X	You might answer questions such as "Which judge will try the case?" "Where?", "What time of the day?"
5.Determine the strategy for a plea for case X	You might weigh the importance of the results of (3) and (4) and take your own capabilities (plea-style) into account when deciding about aspects to include in your plea.
6.Determine the way to proceed from pleastrategy to pleanote to plea in case X	You might write a concept plea-note in a certain format using results of (3) and (5) in spoken language, always keeping your goal in mind and using a certain style to express yourself.
7.Determine the way to proceed from pleanote to plea in case X	You might transform the plea-note into index cards containing the basic line in your plea and then practice this for yourself paying attention to various presentation aspects (verbal and non-verbal behavior).
8. Practice the plea for case X	You might ask friends to give feedback, and can record your own attempts for self evaluation.
9. Plead case X in court	You might pay attention to the reactions of the various listeners.

Tasks in MmPs at the Open University of the Netherlands typically have a study load of about 10 to 20 hours. Despite the interrelatedness of constituent skills, they are easy enough to be dealt with using a whole-task approach. The available instruction time (100-200 hours) is considered enough to master the task on a *basic* level of the professional standard while offering a variability of practice. Schemas for recurrent aspects of the skill are not automated at this basic level. More skilled performance, eventually leading to expertise, is supposed to add at least a factor of ten to the required training time (Eraut, 1997).

An Instructional Design model for MmP-development

Because MmPs focus on non-recurrent aspects of goal competencies using a whole-task approach, traditional ID theories using behavioral task analysis, which is restricted to the analysis of recurrent skills, have several design shortcomings (see, e.g., Dehoney, 1995). Cognitive task analysis is more appropriate here because in focusing on the whole task it can deal with non-recurrent aspects of a complex skill (Merrill, 1987; Reigeluth, 1983; Reigeluth & Merrill, 1984; Tennyson, Elmore, & Snyder, 1991; van Merriënboer, 1997). Furthermore, it also allows the description of expert performance in complex problem-solving domains (Dehoney, 1995; Dubois, Shalin, Levi, & Borman, 1995; Gardner, 1985; Roth & Woods, 1989). It does so by illuminating the *covert heuristics* (Wilson & Cole, 1990) used by experts to solve problems.

An example of an ID-theory focusing on the whole task and herein dealing with the analysis of non-recurrent aspects of complex skills, as well as the teaching of heuristics or rules-of-thumb that help learners to perform such skills, is the four-component ID (4C/ID) model (van Merriënboer, 1997). The 4C/ID-model prescribes task-analytical techniques and design principles for four interrelated components: (a) whole learning tasks, which provide the backbone of any training program for complex learning; (b) supportive information, which helps learners to learn non-recurrent aspects of whole learning tasks; (c) just-in-time information, which is prerequisite to learning recurrent aspects of whole learning tasks; and

(d) part-task practice, which may provide additional practice in recurrent task aspects. The model presented in this chapter is best seen as a specification of the first component, that is, whole learning tasks and related support structures in MmPs.

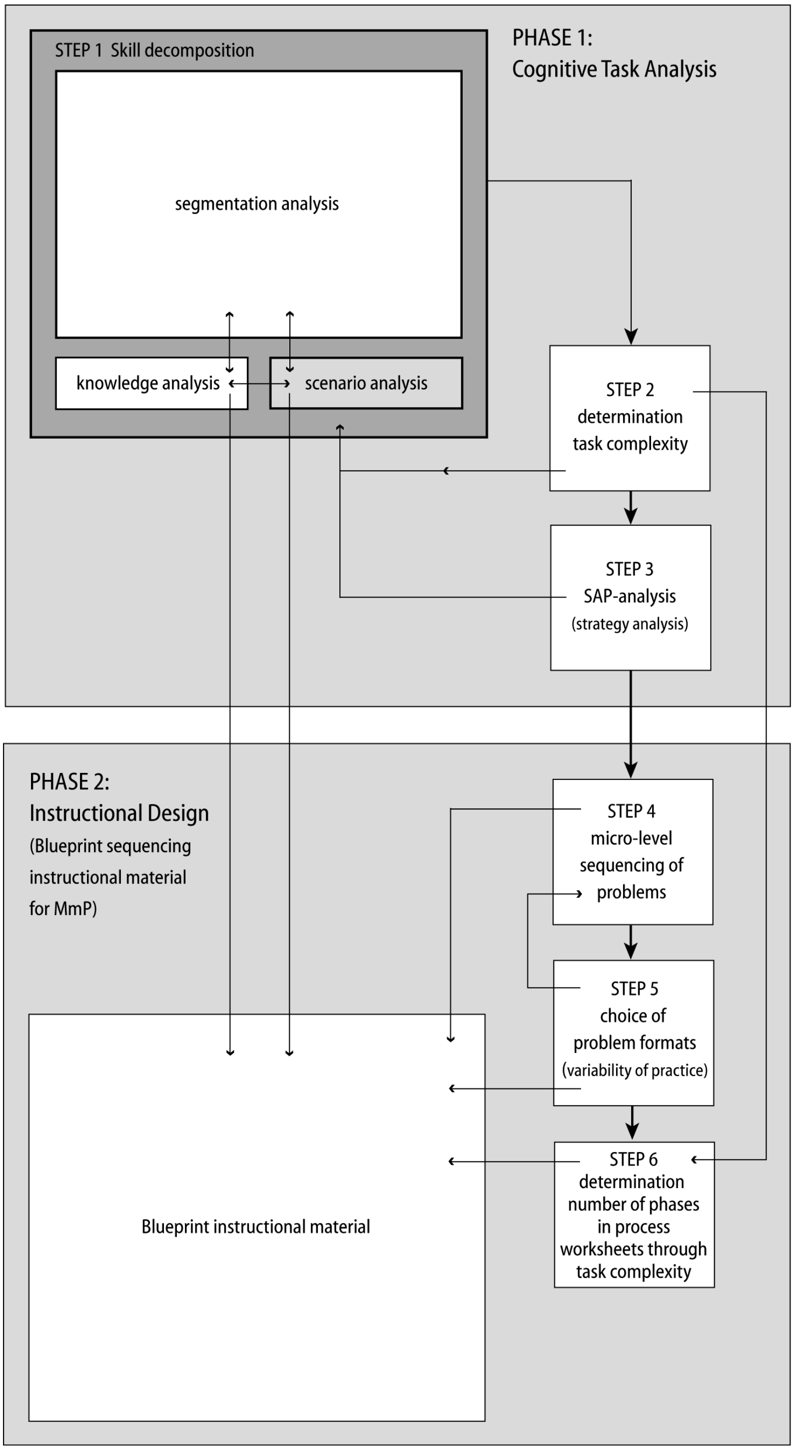


Figure 2.2
The two-phase six-step ID-model for MmP-development.

The two-phase six-step ID-model (Figure 2.2) deals with cognitive task analysis issues (Phase 1: Steps 1 through 3) and ID issues (Phase 2: Steps 4 through 6). It results in a detailed blueprint for the instructional material.

The iterative steps in the model are:

Analysis phase

- Step 1: Skill decomposition based upon task complexity through:
  - segmentation analysis
  - knowledge analysis
  - scenario analysis for identification of problem (in)dependent features
- Step 2: Objective determination of task complexity
- Step 3: SAP-analysis or strategy analysis

Design phase

- Step 4: Micro-level sequencing of problems (inductive)
- Step 5: Choice of problem formats for variability of practice (within the micro-level sequencing)
- Step 6: Choice of the number of phases in the strategic approach to problem solving to be presented to learners via process worksheets

In the following sections the ID activities within the separate steps are described. Since the model was applied in the development of the MmP *Preparing a plea*, examples for clarifying those six steps are taken primarily from it.

Phase 1: Cognitive Task Analysis for MmP-development

Phase 1, consisting of three steps, makes extensive use of experts and focuses on SAP-analysis since this is an important input for ID activities. The steps in the cognitive task analysis are iterative. Skill decomposition (Step 1) identifies segments in a so-called segmentation analysis. Supportive knowledge (resulting from knowledge analysis), strategic knowledge (resulting from SAP-analysis) and more or less problem-dependent features (resulting from scenario analysis) are identified for each segment. SAP-analysis or strategy analysis (Step 3) specifies the time-relationships between the segments (i.e., the constituent skills). As an overarching tactic task complexity is measured (Step 2) to guide the level of detail in the analyses.

Three different categories of experts are used for the different analyses. The first category is practitioners in the problem domain with a lot of experience (here lawyers with more than 10 years of experience: *nestors*). The second category is practitioners new in the domain, but who function as trainers in this domain (here fairly recent graduates who are practicing their profession: *trainers*). The final category is teachers used to teaching in the problem domain, but who no longer practice (*teachers*). Roth and Woods (1989) indicate that the choice of experts is a potential area of bias in the cognitive task analysis. This bias is avoided by using a reasonable amount of experts with different backgrounds. They provide input for various analyses through standardized interviews that are analyzed by instructional designers. In their analyses they look for consensus while identifying (reasons for) observed differences.

Step 1: Skill decomposition

In Step 1, the complex skill for the MmP is decomposed and analyzed. Three types of analysis are used here, namely (a) segmentation analysis for determining relevant segments, (b) knowledge analysis for determining relevant prior knowledge, and (c) scenario analysis for determining possible problem situations for the ID phase. These analyses are conducted, for the most part, in parallel and make use of data generated during Steps 2 and 3 (see Figure 2.3).

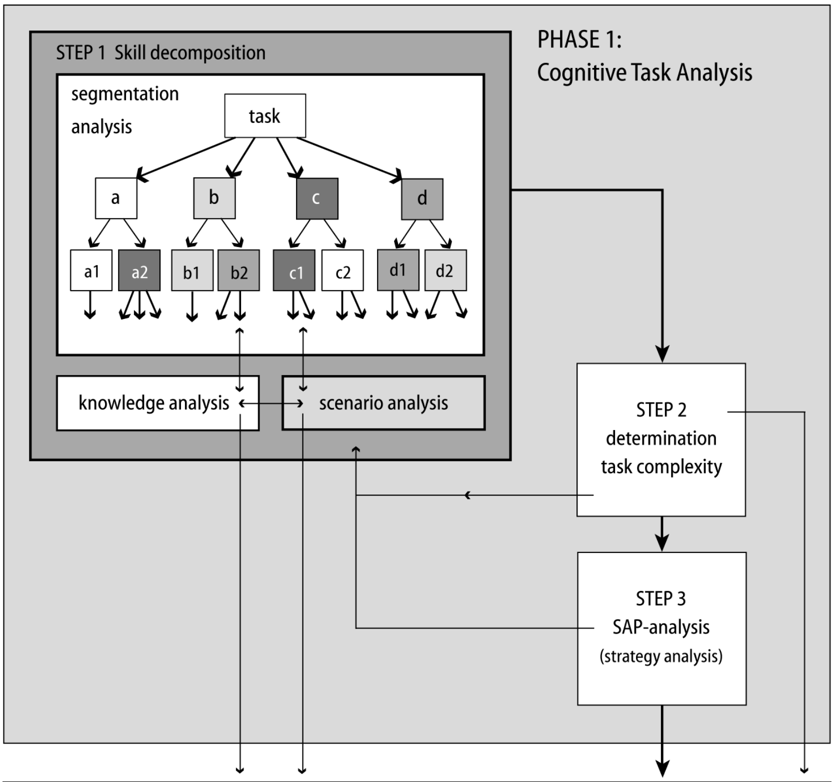


Figure 2.3
Cognitive Task Analysis phase (phase 1) of the model.

Segmentation analysis.
Segmentation analysis results in segments (subtasks, subsubtasks, etc.) of differing sizes. These segments must be both functional and non-trivial and must be of comparable complexity (complexity falls within a predetermined range). This determination makes use of an objective measure of task complexity to optimize task decomposition; that is, preventing too much or too little decomposition. Functional segments enable learners to build relevant schemas while non-trivial segments challenge them (cf. Clark, 1999). Using *teachers* (former practitioners) in this analysis prevents dysfunctional, trivial and/or too complex segments because teachers have the pedagogical experience needed to determine this. For the goal competency, pleading a case in court, the task "order the documents of file X" can be further decomposed into (a) identify legal documents, (b) identify letters, (c) identify notes, and (d) order all three categories chronologically. The teachers identified this as trivial and therefore not to be included in the segmentation analysis. Since teachers use intuitive measures of task complexity to determine when to stop decomposition a task-complexity measurement instrument was needed to be developed (see Step 2) for designing self-contained MmPs.

Knowledge analysis.
Trainers and nestors identify - among other things - *supportive knowledge*; the declarative knowledge that supports the performance of non-recurrent aspects of a skill (van Merriënboer, 1997). Supportive knowledge refers primarily to complex cognitive schemas such as conceptual models, goal-plan models, causal models, and functional models. Supportive knowledge for conducting a plea are for example models that link the consequences of the characteristics of certain plea-styles to actual plea-performance, link the impact of certain (non)verbal behavior on the way people react and describe the ways in which we can attract a person's attention. Supportive knowledge has a bidirectional relationship with strategic

knowledge (which may be either procedurally or declaratively encoded in memory) in supporting the non-recurrent aspects of a skill.

Scenario analysis.

MmPs deal with situations that differ from each other, but that also have elements in common. Scenario analysis identifies these problem-dependent features for use in the instructional design phase; that is, when does a lawyer do this and when, that? All three categories of experts are involved in this analysis. In the example of pleading a case in court this entails a chronological and detailed description of how the plea was prepared and conducted. This task-analytical information guides the process of finding and describing problems or examples and enables designers to design and order problem-situations in the ID phase.

Step 2: Objective determination of task complexity

In Step 2, the complexity of the tasks described in Step 1 is determined using a task-complexity measurement instrument to prevent using tasks that are either too complex or too simple. In this way, learners can be optimally challenged during their learning experience.

Task complexity can be objectively determined (Bonner, 1994; Campbell, 1988; Campbell & Gingrich, 1986; Wood, 1986). To determine task complexity objectively, *teachers* use a task-complexity measurement instrument developed according to the Burtch, Lipscomp, and Wissman procedure (1982). Burtch et al. used a benchmark scaling technique in which anchor tasks described each complexity level on a scale. This technique is easy to use and results in an instrument that can be quickly used for analogous tasks (i.e., domain-specific tasks). Expected prior knowledge of learners is stated in advance of the complexity determination. The effective use of experts in determining task complexity has been reported in various studies (Bonner, 1994; Burtch et al., 1982; Byström & Jarvelin, 1995). Task complexity has proven to be both an effective predictor of task performance (see Bonner, 1994) and a relevant indicator of development costs.

Step 3: SAP analysis

Step 3 of this cognitive task analysis identifies a domain-specific problem-solving strategy together with its associated heuristics. *Trainers* play a key role in identifying this SAP since they themselves, as beginning practitioners, are not far removed from the target population. Their SAPs, acquired through thinking-aloud protocols, can with relatively small changes be used for ID purposes. Nestors internalize, automate, and/or shorten their SAPs to such an extent that they leave out many steps, making it almost impossible to use them for instructional purposes. Practicing law is quite different from learning to practice law (see also Kirschner, 1991, in the domain of the natural sciences). Trainers have not yet internalized, automated, and/or shortened their SAPs to the level that nestors have. An example of a SAP with its related heuristics for *Preparing a plea* (Wöretshofer et al., 2000) has already been illustrated in Table 2.1.

A second problem with using the SAPs provided by experts (Kirschner, 1991) is that the way an expert works in his/her domain (epistemology) is not equivalent to the way one learns in that area (pedagogy). A similar line of reasoning is followed by Dehoney (1995), who reasons that the mental models and strategies of experts have been developed through the slow process of accumulating experience in their domain areas. It is therefore not clear what happens if these models and strategies are imposed on learners. They may interfere in as yet unknown ways with the process of acquiring expertise. Dehoney (1995, p. 120) however proposed that: "some lower-level cognitive strategies can be taught. For example, experts' domain specific strategies for planning and reflecting on the problem solving process will emerge from a cognitive task analysis. These can be taught to novices through modeling". In the research presented here, providing a domain-specific strategy in problem-solving through a process worksheet supports the process of acquiring expertise, because this is an example of such a domain-specific planning strategy.

For achieving goal competency using a whole-task instead of a part-task approach is advocated because the learner quickly acquires a view of the whole skill (Reigeluth, 1987). A second advantage of a whole-task approach with a trainer SAP is that learners can use the output from one task assignment as input for the following. In other words, the task is more authentic. Finally, whole-task practice aims at inductive processing in which complex cognitive skills are acquired by practicing them under different conditions (e.g. different problem formats, different sequencing principles, and fading of scaffolding). In this approach, *induction* of cognitive schema is promoted by concrete experiences that force the learner to work from given examples to more general and abstract knowledge and strategies. For example, in *Preparing a plea* (Wöretshofer et al., 2000) important elements are presented via a "virtual video tape" containing examples and nonexamples of certain plea behaviors. Each concrete observable behavior in a plea is directed at the achievement of a certain subgoal (e.g., make information accessible, keep someone's attention)

Phase 2: Designing instruction for MmPs

Phase 2 uses the results from the cognitive task analysis-phase and focuses on micro-level sequencing of the tasks (Step 4), choosing relevant problem formats (Step 5) and choosing the appropriate number of phases for process worksheets (Step 6). These steps are also conducted iteratively and result in a detailed blueprint for the MmP (see Figure 2.4). This leads to micro-level high-variability sequencing using worked-out examples and problems with process worksheets (with a certain number or phases for SAPs included). This approach both encourages schema construction and supports transfer, and corresponds with the earlier enumerated guidelines in the 4C/ID model.

Sequencing of learning tasks (Step 4) is pivotal in facilitating the learning process (e.g., Brown et al., 1989; Gagné et al., 1993; Merrill, 1987; Reigeluth, 1983). Although many design methods deal with sequencing instruction, few deal with doing this for complex cognitive skills (see van Merriënboer, 1997).

Working memory is limited. Since learning tasks differ in their taxing of the learner’s working memory, cognitive load theory (Sweller, 1988) was chosen to guide the selection of problem formats (Step 5).

Process worksheets guide learners through the application of the (sub)SAPs needed for performing the task. The optimal number of phases in such process worksheets is determined (Step 6) through determining the task complexity.

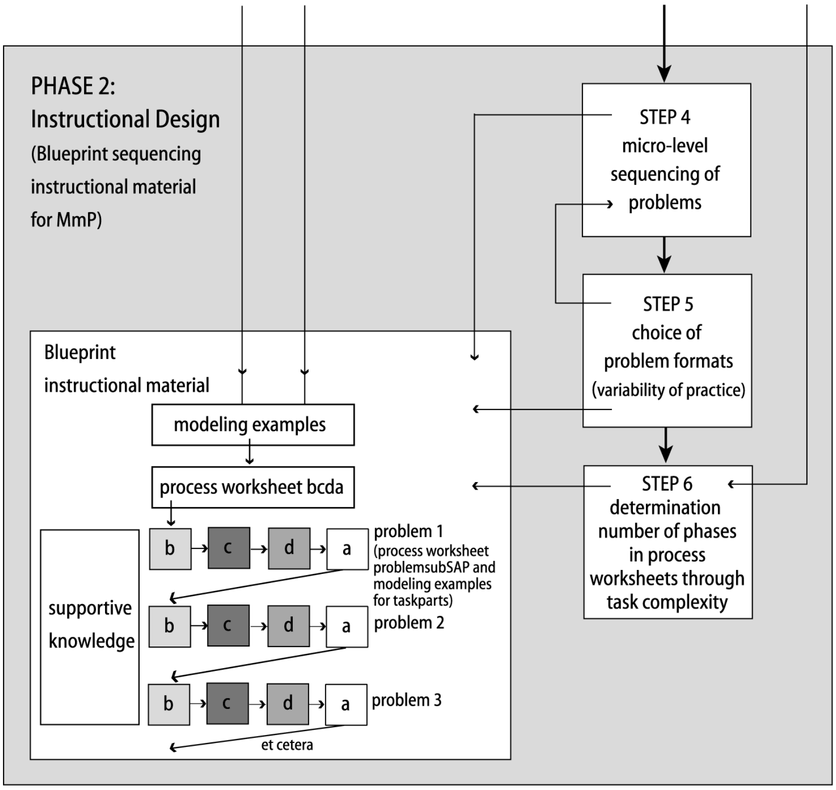


Figure 2.4
Instructional Design phase (phase 2) of the model.

Step 4: Micro-level sequencing

Micro-level sequencing establishes the order in which different problem formats will occur. Typical simple-to-complex ordering of conventional problems is not the most effective approach to micro-level sequencing of the whole task because it tends to hamper schema acquisition. The interconnectedness of the various part-tasks is high which results in increased cognitive load for the learner. High-variability sequencing, on the other hand, provokes inductive processing and improves transfer of training (Spiro, Coulson, Feltovich, & Anderson, 1988).

Nestors, teachers and trainers identify salient features of problems during the scenario analysis (Step 1). Varying tasks and practice with respect to problem situations or conditions (presentation mode, saliency of defining characteristics, task performance contexts) encourages learners to develop meaningful schemas by increasing both the chances that similar features are identified and the chances that relevant features can be distinguished from irrelevant ones. This consistently results in beneficial effects on transfer of training (Cormier & Hagman, 1987; Shapiro & Schmidt, 1982; Singley & Anderson, 1989). A negative aspect is that variability of practice also increases cognitive load. This disadvantage is outweighed by the fact that the alternative - simple-to-complex ordering - seldom has beneficial effects on transfer.

Step 5: Choosing problem formats

Problem formats are used that avoid cognitive overload. Cognitive load theory (Sweller, 1988; Sweller & Chandler, 1994; Sweller, van Merriënboer, & Paas, 1998) can be used to guide the selection of problem formats. Cognitive load theory, with respect to schema learning, prescribes that instruction should be designed such that working memory is capable of processing the instruction. Appropriate problem formats for schema acquisition are (a) real-life conventional problems, (b) product-oriented problems such as worked-out examples and completion problems where learners have to complete a partially given solution, or (c) process-oriented problems such as modeling examples and problems with performance constraints (i.e., process support problems). This theory can be summarized in two basic principles, namely prevent cognitive overload, and redirect attention. Preventing cognitive overload entails posing problems that are not significantly beyond the learner's level of competence. Redirecting attention shifts learner attention from cognitive processes not directly relevant for learning (e.g., searching information, weak-method problem solving) to processes relevant for learning (in particular, schema construction by induction from concrete cases). Sweller's approach through using different problem formats and fading support as the learner gains more expertise was augmented. Of the problem formats suitable for achieving non-recurrent skills, worked-out examples and problems with performance constraints combined with process worksheets typically meet the criterion of preventing cognitive overload (van Merriënboer, 1997, p. 187).

Learners using these MmPs receive a process worksheet based on a domain-specific problem solving strategy (SAP) to solve the problem tasks presented. This approach is beneficial in that it encourages the development of schemas (Bedard & Chi, 1992; Gagné et al., 1993; Sternberg & Frensch, 1991).

Cognitive load theory predicts an interaction between problem formats and sequencing (Paas & van Merriënboer, 1994). For problem formats with high cognitive load, changes in variability of the sequence have little or no effect on inductive processing and transfer due to the possibility of cognitive overload. For problem formats with relatively low cognitive load, increasing the variability of the problem sequence will substantially enhance inductive processing and transfer (Paas & van Merriënboer, 1994; van Merriënboer, Schuurman, de Croock, & Paas, 2002). High variability sequencing may have positive effects on transfer, but it also has negative effects on the number of problems or training time needed to reach a certain performance level. Thus, in the same training time, fewer problems can be solved; otherwise, more training time is needed to reach a predefined performance level. This is called the *transfer paradox* (de Croock, 1999; Jelsma, 1989; van Merriënboer, de Croock, & Jelsma, 1997).

Table 2.2 contains the sequence and problem formats constituting the blueprint that was identified for the example, *Preparing a plea* (Wöretshofer et al., 2000).

Table 2.2
Blueprint of instruction within Multimedia Practical *Preparing a plea*

1. Modeling example(s) (video-registration of persons conducting a plea)	
2. Presentation of process worksheet containing phases 1-8 to proceed from file to plea	
3. File Bosmans (civil law) (Problem 1 consisting of *i* assignments)	
TA 1	Order documents in categories (practice-files for Phase 1) (problem with process worksheet subSAP for Phase 1)
TA 2	Get acquainted with file using guiding questions (practice-files for Phase 2) (problem with process worksheet subSAP for Phase 2, including worked example for Phase 1)
TA *i*	Task assignment *i* (practice-files for Phase *i*)(problem with process worksheet subSAP for Phase *i*, including worked example for Phase *i*-1)
4. File Ter Zijde (criminal law) (Problem 2 consisting of *i* assignments)	
(No practice files and less in-phase cueing as compared to File Bosmans; i.e., fading)	
TA 1	Order documents in categories (problem with process worksheet subSAP for Phase 1)
TA 2	Get acquainted with file using guiding questions (problem with process worksheet subSAP for Phase 2)
TA *i*	Task assignment *i* (problem with process worksheet subSAP for Phase *i*)

SAP = systematic approach to problem solving TA= Task assignment

The design prescription used here entails first designing whole-task practice aimed at inductive processing through using a high variability sequence, followed by fading as learners gain more expertise, proceeding from concrete modeling examples to problems with process worksheets to enhance inductive processing and transfer. This is a specification of the box "Blueprint instructional material" in Figure 2.4.

Step 6: Determination of number of phases within process worksheets

The final step - before actual development of the instructional material - is determining the number of phases within the process worksheet, which will guide them through the problem solving process of the whole task. The complexity of the task assignments and the prior knowledge of the learners primarily influence the number of phases.

Determining the number of phases is an optimization problem (Clark, 1999). The amount of work in ID and development activities (and thus also the cost) is directly proportional to the number of phases. The more phases, the higher the cost. Instructional design theorists agree that *designers* should decompose tasks in the analysis phase to a greater level of detail than that which is presented to the *learner* (Jonassen & Hannum, 1995; Jonassen, Hannum, & Tessmer, 1989, 1999; Merrill, 1983). However, little is known about what size of segments and herein what number of phases should be used within a process worksheet for learners with a particular level of prior knowledge, given a certain task decomposition and a specified task complexity. As soon as content and learning goals are determined, the optimal number of segments is mainly influenced by prior knowledge and skills (Chang, Ho, & Liao, 1997; Kalyuga, Chandler, & Sweller, 1998).

In the analysis phase it was stressed the importance of determining objective task complexity for task decomposition and identification of a SAP by trainers. This task complexity also guides the process of determining the number of phases within process worksheets in MmPs.

Tasks used in instruction should preferably have sufficient and comparable complexity to challenge learner capacity. Too detailed a decomposition results in tasks that are too simple. Too global a decomposition results in tasks that are too difficult. Suppose that trainers identify a SAP referring to the tasks "bcda" during a cognitive task analysis, where each task assignment can further decomposed into smaller task assignments (task assignments a, b, c, d, a_1, a_2, b_1, b_2 et cetera). Theoretically, then, "bcda" can be presented in a process worksheet as $(b_1b_2)(c_1c_2)(d_1d_2)(a_1a_2)$ or $b(c_1c_2)da$ or $(b_1b_2)c(d_1d_2)(a_1a_2)$ et cetera (see Figure 2.3). For each of those task assignments, task complexity needs to be determined using an instrument for measuring task complexity. In the example in Figure 2.4 it has been decided to present "bcda" (not further decomposed) in a process worksheet since the task assignments "a", "b", "c" and "d" have sufficient and comparable complexity.

First experiences with the model in developing the MmP, *Preparing a plea*

In the development of *Preparing a plea* (see Appendix 1, Wöretshofer et al., 2000) steps 1, 3, 4 and 5 of the two-phase six-step model were applied. The instrument for objectively measuring task complexity (Steps 2 and 6) was not yet available during development, but was tested afterwards. This developmental study was intended to determine (a) if an objective task complexity measurement really was necessary, and (b) the instructional effectiveness of the material developed according to the model.

Method

Participants, materials and procedure

Twenty experts from three different backgrounds participated in the cognitive task analysis: eight practicing lawyers (‘nestors’), six trainers of starting lawyers who also practice as lawyers themselves (‘trainers’), and six law teachers familiar with teaching students to conduct a plea (‘teachers’). The structured interview technique was used in order to get an impression of experts’ ideas about how to prepare a plea (cf. Cooke, 1994).

The MmP *Preparing a plea* (Wöretshofer et al., 2000), was developed according to the instructional blueprint resulting from Steps 4 through 6. The six teachers were asked to *subjectively* determine task complexity on a 4-point rating scale (i.e., without anchor tasks). Together with three instructional developers, those law teachers were also involved in the actual development of the MmP. After development, a second group of 32 participants was asked to *objectively* determine task complexity using a task-complexity measurement instrument. This instrument was developed according to the procedure of Burtch et al. (1982), and used a benchmark scale with four anchor tasks, each describing one complexity level (see Chapter 3 of this thesis).

The MmP was studied by a small group of sophomore law students ($N = 12$). Six of them had no plea experience at all; the other six had some limited plea experience as members of a debating club. The MmP was developed for use by students without plea experience. A jury consisting of three persons (two teachers and one trainer) scored students' results on the pleas. Data were collected on subjective perception of task complexity, student motivation, confidence to plea without a process worksheet, and appropriateness of the number of phases.

Results and discussion

The qualitative data gathered in the structured interviews demonstrated an interesting difference between trainers' and nestors' SAPs, namely that trainers gave more elaborate descriptions of how to prepare a plea and thus probably had not internalized and automated SAPs to the level nestors have. The development group of six law teachers was asked to *subjectively* determine task complexity for the nine task assignments to be included in *Preparing a plea* (Wöretshofer et al., 2000) on a 4-point rating scale. The subjective task complexity of the nine task assignments ranged from $M = 1.2$ (for the task "ordering the file", $SD = .4$) to $M = 3.3$ (for "setting up a plea strategy", $SD = .8$). A satisfactory interobserver agreement for the judged complexity of tasks was found, Kendall’s $W = .53$, $X^2 = 25.56$, $p < .01$.

In the retrospective analysis of task complexity by the second group of thirty-two teachers, the judged objective task complexity of the nine task assignments ranged from $M = 1.3$ (for the task "ordering the file", $SD = .7$) to $M = 3.5$ (for "setting up a plea strategy", $SD = .8$). Again, a satisfactory interobserver agreement for the judged complexity of tasks was found, Kendall’s $W = .32$, $X^2 = 50.59$, $p < .01$. Whereas the subjective and objective judgments show agreement on what the simplest task is ("ordering the file") and what the most complex task is ("setting up a plea strategy"), there is a difference in the overall rating of task complexity. The mean subjective complexity over the nine task assignments ($M = 2.4$, $SD = .3$) is significantly higher than the mean objective complexity ($M = 2.1$, $SD = .4$; $t(24) = 2.1$, $p < .05$). In other words, if objective instead of subjective measures had been used during the MmP development this would have yielded a smaller number of phases.

All 12 sophomore Law students successfully completed the MmP and were, according to the jury, able to conduct a plea in court. Table 2.3 presents the results on subjective task complexity, motivation, confidence, and appropriateness of the number of phases for the subgroups with no plea experience and with limited plea experience. Students reported a mean subjective task complexity of 3.45 on a 9-point scale, indicating a rather low subjective complexity; a mean motivation of 3.38 on a 4-point scale, indicating a high motivation, and a mean confidence in their ability to plea without a process worksheet of 2.28 on a 4-point scale, indicating moderate confidence. There were no significant differences between subgroups for subjective task complexity, motivation and confidence. However, with respect to the appropriateness of the number of phases, students without plea experience agreed significantly more with the number of phases ($M = 2.6$, $SD = .5$) than students with limited experience, who reported that the number of phases could be fewer ($M = 1.72$, $SD = .75$; $t(10) = 4.94$, $p < .05$).

Table 2.3
Learners' results with Multimedia Practical (MmP) *Preparing a plea*

	No Plea Experience (n = 6)		*Limited Plea Experience (n = 6)*	
Measure	*M*	*SD*	*M*	*SD*
Subjective Task Complexity [1-9 (very, very complex)]	3.7	.8	3.2	.8
Motivation [1-4 (very high)]	3.5	.5	3.25	.7
Confidence [1-4 (very high)]	1.8	1.0	2.75	1.0
Too few phases [1-4 (completely agree)]	2.6	.5	1.72*	.8

* *p* < .05

Concluding, the subjective task complexity data of both teachers and students show that the task assignments were not too difficult. The objective task complexity measured by teachers indicates that the number of phases could have been somewhat less, especially for those students with some prior plea experience. Students reported high motivation, confirming that the number of phases was not too high and that task assignments were experienced as functional and non-trivial. The findings show that the instructional materials developed according to the model are effective and they give tentative support for the use of an objective task-complexity measurement instrument in ID.

General discussion

The two-phase six-step model has proven to result in a detailed blueprint for effective MmP development. The cognitive task analysis phase results in detailed input for the design phase that is largely based on the 4C/ID-model (van Merriënboer, 1997) and insights from cognitive load theory. Objective measurement of task complexity determines the optimal number of phases for the SAP through a process worksheet to be presented to learners.

The benefit of cognitive task analysis is clearly its rich and thorough description of task performance. This approach, however, is not without drawbacks. The largest problem is resource intensity for both data gathering and data analysis. A second problem is that it is susceptible to bias and error and should be used by experienced instructional designers because it is mainly heuristic in nature (Jonassen & Hannum, 1995; Jonassen, Tessmer, & Hannum, 1999). In fact, the model has been described in this chapter by giving a SAP and associated rules-of-thumb or heuristics for its use. More research and development should be directed towards further specification and articulation of the model, which is necessary to make it directly useful for less experienced designers or teachers. This is in line with formative research on the simplifying conditions method (Reigeluth, Lee, Peterson, & Chavez, 1999). At this moment, however, no other less costly and less error prone methods are available.

In defense of the model, we are convinced that it will save costs on ID and development and result in reusable, high-quality materials. The present costs for designing and developing computer assisted multimedia instruction greatly exceeds the costs incurred by cognitive task analysis. Beyond this, there is preliminary evidence that the use of various types of experts (nestors, trainers and teachers) also prevents serious mistakes in the analysis and design phases, especially if this is accompanied by an instrument for determining task complexity. In this, the model clearly differs from other task-analytical methods which almost always include the use of one particular category of experts and which do not consider measuring task complexity. An important critical success factor for developing instruction is that the instructional designer needs, to a certain extent, to be familiar with the subject domain (Dehoney, 1995).

Because the model has not yet been broadly applied, research should be conducted to further justify our assertions. The most important question to be addressed at this moment is that of the number of phases within process worksheets by using a task-complexity measurement instrument. As far as we know, no studies examine the impact of the number of phases on task performance and efficiency. Through a study on process support in which the number of phases is varied it is expected to gain more insight into this matter. This study is described in Chapter 4 of this thesis. Before turning to this study, we will first describe the development of the instrument for objectively determining task complexity in Chapter 3.

Chapter 3 - Development of an instrument for measuring the complexity of learning tasks*

An instrument for measuring the complexity of learning tasks in the field of Law was developed and tested in three experiments. In Experiments 1 and 2, teachers used the card-sort method to rate the complexity of learning tasks. Based on the outcomes, a benchmark scale with four criterion tasks was used in Experiment 3. The results showed the benchmark instrument to be valid and easy to use, allowing instructional designers to design competency-based learning environments that better take task complexity into account.

Introduction

As was apparent in the previous chapter, determining task complexity is necessary for achieving optimal decomposition of learning tasks. Task complexity has both an objective and a subjective component. Objective task complexity results from the characteristics or the nature of the task itself. Subjective complexity is determined by the characteristics of the task and of the person carrying out the task. Playing an etude from Chopin, for example, is objectively more complex than practicing the scales on a piano. This is 'objectively' true for both the expert and the novice, although the expert will 'subjectively' experience playing Chopin as being less complex than the novice will. In a more cognitive vein, sentence complexity is another example of where complexity can be objectively determined, irrespective of the readers' familiarity with the content of the sentence. Most often, sentence complexity and readability are determined on the basis of sentence length, word length, number of phrases and clauses, et cetera (e.g., Flesch, 2003; Vaso, 2000). Although such readability formulas are not undisputed (Brandle, 2002; Clough, 2000; Pikulski, 2002), one would agree that a sentence in which the subject and object are separated by a large number of dependent and independent clauses and where the average word length is quite long is more complex and thus more difficult to understand than a simple sentence (see this sentence!). Again, the experienced reader will have an easier time than the novice (subjective), but this does not nullify the fact that sentences also objectively differ.

Tasks that consist of higher-level unique constituent skills requiring more coordination have higher objective complexity than tasks with fewer unique constituent skills requiring less coordination. Subjective task complexity is the complexity experienced by the learners while performing the task as a reaction to the task characteristics, their own characteristics, and the characteristics of the environment.

Studies have shown that task complexity can be used to predict task performance. This is true for both objective task complexity (e.g., Boggs & Simon, 1968; Early, 1985; Kernan, Bruning, & Miller-Guhde, 1994; Scott, Fahr, & Podsakoff, 1988) and subjective task complexity (e.g., Huber, 1985; Taylor, 1981). While these studies focused on either objective or subjective task complexity, a more recent study by Maynard and Hakel (1997) explicitly focused on uncovering the relationships between the two. What they found was that objective task complexity is a good predictor of subjective task complexity, in the sense that higher

---

* Based on: Nadolski, R. J., Kirschner, P. A., van Merriënboer, J. J. G., & Wöretshofer, J. (in press). Development of an instrument for measuring the complexity of learning tasks. *Educational Research and Evaluation.*

levels of objective task complexity lead to higher levels of subjective task complexity. In addition, their research showed a high correlation between perceived (subjective) and objective task complexity, a finding consistent with results from earlier studies (Huber, 1985; Kernan et al., 1994; Scott et al., 1988).

The present study concerns the development of a reliable, valid and easy to use measurement instrument for rating the objective complexity of Law learning tasks. Several domain-independent instruments have been developed to determine objective task complexity (e.g., Bonner, 1994; Byström & Järvelin, 1995; Campbell, 1988; Campbell & Gingrich, 1986; Wood, 1986). The main problem with these instruments is that they are difficult to use and usually involve considerable training. Wood (1986), for example, has developed an instrument that makes total task complexity operational by distinguishing between component complexity, coordinative complexity, and dynamic complexity of a task. Component complexity is a direct function of the number of distinct acts executed in the performance of the task and the number of distinct information cues processed in the performance of those acts. Coordinative complexity refers to the nature of the relationships between task inputs and task products, the nature of the relationship is given between 'n' task input(s) and 'm' task output(s) ($n = 1, 2\ldots..$; $m = 1, 2\ldots\ldots$). Dynamic complexity refers to how often individuals must adapt to changes in the cause-effect chain or in the means-ends hierarchy for a task *during* the performance of a task, due to changes in the world which have an effect on the relationship between task inputs and products. For instance, a pilot when landing a plane has to respond to changing weather conditions, the height above sea level and the height above the landing strip. Application of Wood's model requires determination of these three types of complexity and weighting factors for each of them in order to finally determine task complexity.

Campbell (1988) has offered an approach that has been shown to have empirical value in the field of business administration curricula. According to him, task complexity is directly related to those task characteristics that increase information load (i.e., the number of dimensions of information requiring attention), information diversity (i.e., the number of alternatives associated with each dimension), and/or the rate of information change (i.e., the degree of uncertainty involved). He identifies four basic dichotomous task characteristics that affect information load, diversity and/or change namely the presence or absence of:
(1) multiple potential ways ("paths") to arrive at a desired end-state; (2) multiple desired outcomes to be attained; (3) conflicting interdependence among paths to multiple outcomes, and (4) uncertain or probabilistic links among paths and outcomes. On the basis of these four characteristics, sixteen task-types can be distinguished (presence/absence of each of the four task characteristics). However, an exact ordering of tasks from simple to complex is difficult because Campbell does not specify the relative contribution or weight of each of the four basic attributes.

Burtch, Lipscomb, and Wissman (1982) described a simpler benchmark scaling technique in which anchor tasks are used to describe each complexity level on a scale. New tasks are compared to the anchor tasks and the best likeness determines the complexity. This is similar to the Mohr-scale for determining the hardness of minerals where a mineral is scratched with the ‘anchor’ minerals for comparison; the harder mineral leaves a scratch on the softer one. Such an instrument requires very little learning and training; subject matter experts in the task domain can easily use the instrument if the expected prior knowledge of task performers or learners has been defined. The research here has applied this general approach for the development of a benchmark instrument using a conceptual frame of reference largely based on Merrill's Component Design Theory (1987).

As was seen from the description of earlier approaches (Wood, 1986; Campbell, 1988) determining weighting factors for the relative contribution of the various attributes to objective complexity is often difficult or even unknown (Campbell, 1988). Since the conceptual frame of reference in this research centers on learning tasks and concerns *intellectual operations* involved in learning it was attempted to alleviate this problem by using Merrill's Component Design Theory (1987) which distinguishes four major *hierarchical* categories of operations ("performances") that can be defined as the four levels of complexity (i.e., very simple, simple, complex, and very complex). All levels are relative to the prior knowledge of the learners because they are based on the unfamiliarity of the learner with the learning task in which this operation occurs. Once the learner has mastered the learning task in question, this same task becomes routine, therefore becoming simpler than it was before. Complexity increases from (1) *remember an instance*: gain and remember facts / retention (very simple); (2) *understand a generality*: gain generalized, abstract knowledge / insight or understanding (simple); (3) *use*: apply knowledge in *familiar* settings (complex); and (4) *find*: apply knowledge in *unfamiliar* settings / problem solving and qualitative reasoning (very complex). The intellectual operations in the frame of reference used in this study are hierarchically ordered with each higher level subsuming the previous ones. But as is the case for the discriminating characteristics in Campbell's model, an exact ordering of complexity remains difficult since the relative contribution of each of the four classes of intellectual operations to a particular learning task is unknown while the breadth of a certain class of intellectual operations can be very large. This means that under certain circumstances understanding a generality (e.g., understanding the concept Justice) can be more complex than using knowledge (e.g., applying a simple procedure for determining the maximum punishment for a certain crime).

This conceptual frame of reference was used for the development of a benchmark instrument for measuring the complexity of Law learning tasks. The development entailed carrying out three related experiments. The next section describes the general methodology of all three experiments.

General methodology

A similar methodology was used for all three experiments. Where relevant, differences are given when separate experiments are described.

Participants

Two groups of participants were used. One group was composed of Law teachers at Dutch universities from the fields of Criminal Law and Civil Law ($n = 33$). The second was composed of graduate level Law students at Dutch universities ($n = 12$).

For the first group, 23 teachers working at different Dutch universities registered before the start of Experiment 1. Ten additional teachers registered while conducting Experiment 1. No participants from the teacher-group participated in both Experiments 2 and 3 since Experiment 3 included tasks from Experiment 2. Teachers participated in two experiments maximally (Experiments 1 and 2 or Experiments 1 and 3). Participants from the student-group ($n = 12$) only took part in Experiment 3.

Material development

The basis material used in this research was taken from existing Law courses, some of which were competency-based.

The - to be rated - Law learning tasks for the various experiments were restricted using two simple guidelines. First, the tasks were suitable for sophomore Law students. Since all Dutch universities have almost identical Law curricula for the freshman year, all sophomore students can be expected to have comparable prior knowledge and thus the Law teachers could be expected to have similar views of what these students should be able to do. Tasks from exotic sub domains of Law were also excluded. Second, the length of the tasks (formulation plus solution in keywords) was standardized so that "length of task" would not be a contaminating artifact in the determination of complexity.

The two guidelines in conjunction with the conceptual frame of reference were used to determine 56 tasks to be included in the various experiments (see Appendix 2). All four members of the development team (two criminal law teachers, one civil law teacher and one educational technologist specialized in Law courses) independently scored the complexity of the tasks on a 4-point ranking scale (very simple, simple, complex, very complex). The conceptual frame of reference, which formed the criteria for determining the complexity, was known to all of them. There was no simple algorithm for applying the frame of reference so it was possible for the developers to apply the criteria differently. For all 56 tasks, Cohen's Kappa was calculated ($K = 1$ for 46 tasks, $K = .7$ for 10 tasks). After rating, the development team discussed their ratings for further articulation of their conceptual frame of reference.

Procedure

All participants were informed about the experiments, the time schedule and the estimated workload. All printed materials (including instructions) were sent to the participants' work addresses. They had ten workdays to return the materials in a stamped self-addressed envelope. They were informed that they should work individually and that it would take them approximately three hours to do the necessary work. Participants were offered the opportunity to receive further information (by mail or phone). In all three experiments no participant made use of this offer. A reminder was sent when the deadline for return had expired. Upon the completion of an experiment, participants were thanked for their participation, received their compensation and were informed about their participation in the upcoming experiments. Participants were compensated with a small gift plus a small monetary remuneration (circa $80) per experiment.

Experiment 1

One important criterion for an easily usable benchmark scale is its non-specificity for raters' area of expertise. Experiment 1 studied whether the specific expertise of a participant in a sub domain of Law (Civil or Criminal) influenced the rating of tasks from their own or from the other sub domain. The experiment was also used to begin the process of determining anchor-tasks for further experiments and as a pilot for the design of the questionnaire to be used in the further experiments.

Method

Participants

Nineteen law teachers (7 Criminal Law, 12 Civil Law) employed at Dutch universities (10 distance education, 9 face-to-face education), returned their results (response rate = 83%).

Materials

The materials consisted of Criminal Law and Civil Law learning tasks in two separate packages plus a series of questionnaires. Each task package contained 16 Law learning tasks selected from the original 56, one task per page. The instrument for gathering the data consisted of seven different parts:

1. Card sort task for complexity. Participants were asked to sort each of the 16 tasks into four *equal* piles with comparable complexity (very simple, simple, complex, very complex). The tasks provided had - according to the development team - an equal distribution within the conceptual frame of reference (i.e., four tasks for each category).
2. Task ranking within piles. Once participants had made the four equal piles, they ranked the tasks within each pile from least to most complex. As a result, for both sub domains, the 16 tasks were sorted with respect to increasing complexity on a 16-point ranking scale.
3. Students' time on task estimations. Participants indicated how long they felt it would take a sophomore to *learn to perform* each task: this 'learn to perform' is stressed as for instance 'to learn to perform a plea' is more time-consuming than 'to perform a plea'. Time to conduct a task is considered by some researchers to be a good indicator of task complexity (Maynard & Hakel, 1997; Winne, 1997).
4. Rating criteria. To determine their conceptual frame of reference for determining task complexity, participants were asked to rate 18 assertions on possible criteria for judging the complexity of Law learning tasks on a 4-point categorical scale ranging from totally disagree (1) to totally agree (4). Assertions dealt with topics such as 'number of possible solutions', 'kind of intellectual operations required', et cetera. There was space left for the participants to add other topics.
5. Participants' time on task. To determine the speed of use of the different instruments, the time needed to carry out the 'card-sort and ranking'-task as well as for 'estimating the students' time on task'-task was reported by participants.
6. Ease of use. Since speed of use is not necessarily the same as ease of use, a 9-point categorical scale developed by Paas and van Merriënboer (1994) was used to measure the perceived cognitive load of the (1) card-sorting task, (2) the ranking task, and (3) the 'estimated students' time on task'-task. Cognitive load is supposed to be an indication for ease of use; the less mentally demanding the task, the lower the cognitive load. This was included to check the perceived cognitive load of what the participants were asked to do and thus to check if the instrument is easy to use.
7. General information. Data were collected on participants' experience, gender, et cetera.

Design and procedure

A 2x2 (expertise x sub domain) completely crossed, factorial design was employed. The expertise of the rater could be in Criminal Law or Civil Law as could be the sub domain of the learning tasks.

Participants were asked to sort the learning tasks provided (formulation plus solution in keywords) with respect to their judgment of the complexity for sophomore Law students to *learn* to carry them out. It must be stressed here that the participants did *not* rate how complex it would be to carry out the task, but rather how complex it is to LEARN how to carry out the task. For example, learning to walk a tightrope is a complex task, whereas once having mastered this, it becomes quite easy for the tightrope walker. Tasks were randomly ordered for the card sort. Task ranking within the four piles and students' time on task estimation followed this.

Results

Participants' expertise

It was expected that the specific field of expertise of participants would not influence their ratings for the sub domains, since all participants had experience with all offered tasks during their own study. In other words, a teacher of Criminal Law would also be familiar enough with sophomore Civil Law learning tasks to rate them with a result similar to the teacher of Civil Law teacher. Secondly, since all freshman law curricula are (almost) identical at all Dutch universities and all faculty members at the Open University of the Netherlands (distance education) are products of "traditional" face-to-face universities, it was expected that the "type of university" of the participant (distance education vs. face-to-face) also would not influence their ratings.

An univariate analysis of variance for the sum of deviations of participants' ratings to the conceptual frame of reference revealed no significant differences in participants' ratings for Criminal Law versus Civil Law learning tasks based upon their area of expertise (Criminal Law tasks, $F(11, 6) = .029$, $MSE = .303$, $p = .867$; Civil Law tasks, $F(11, 6) = .004$, $MSE = .063$, $p = .948$). For all tasks taken together the area of the participants’ expertise did not influence their ratings for Criminal Law learning tasks and Civil Law learning tasks. The results for all separate tasks also showed the same pattern. Participants also indicated that they did not expect themselves to rate tasks in their own sub domain of expertise better, as confirmed by the rating results (cf. self-efficacy: Bandura, 1982). Participants regarded their 'teaching expertise' of slightly more importance than their 'subject matter expertise' for the quality of their ratings. This difference, however, was not significant.

A second univariate analysis of variance for the sum of deviations of participants' ratings to the conceptual frame of reference showed that "type of university" also did not affect ratings for both groups of tasks as a whole (Criminal Law tasks, $F(12, 6) = .067$, $MSE = .936$, $p = .799$; Civil Law tasks, $F(12, 6) = .038$, $MSE = .395$, $p = .848$) nor for all separate tasks.

Card sort

To estimate the extent to which the individual ratings of the participants in the card-sort tasks correspond with each other, the concordance coefficient - Kendall’s *W* - was calculated (Hays, 1981; Siegel, 1956). This coefficient was calculated for all four conditions and for both the 16-points ranking scale and 4-points ranking scale (Table 3.1).

Table 3.1
Concordance coefficient for criminal and civil law tasks by participants' area of expertise

	Criminal law tasks		Civil law tasks	
Participants' expertise	4-points	16-points	4-points	16-points
Criminal law (n = 7)	.547	.570	.518	.592
Civil law (n = 12)	.601	.651	.540	.613

p < .01 for all values for Kendalls W (concordance-coefficient)

All coefficients are significant at the 1% level of probability confirming that the participants showed a large degree of agreement on the rankings and ratings and that the participants were applying the same standard in ranking the tasks under study.

Table 3.2 presents the descriptive statistics for the separate tasks in the card sort and estimated student's time on task; the latter can be disregarded for the moment. The order of the tasks in the card sort from very simple to very complex was determined by the mean rating scores. The classification of a task in one of the four categories on the basis of the mean score or on the basis of the median is the same for all tasks. The data showed that participants' ratings for the separate tasks differed quite a lot. If differences occurred between raters' classifications and the conceptual frame of reference, the deviation was maximally one class. The data presented in Table 3.2 show that the consensus among participants for the extremes (very simple tasks and very complex tasks) was larger than for the two intermediate categories. The rating-values based on the median (*Mdn*) and the values based on the conceptual frame of reference (Rf) were much more in correspondence for both very simple tasks and very complex tasks than for the other two categories.

Table 3.2
Ratings for criminal and civil law learning tasks (4-points-ranking scale) and students' time on task statistics

Criminal law tasks									Civil law tasks								
card sort						time on task			card sort						time on task		
Id	*M*	*SD*	*Mdn*	Rm, Rf	P(c = ci)	*M*	*SD*	rm	Id	*M*	*SD*	*Mdn*	Rm, Rf	P(c = ci)	*M*	*SD*	rm
cr14	1.32	.58	c1	c1, c1	.72	13	8	c1	ci8	1.21	.42	c1	c1, c1	.86	11	5	c1
cr8	1.37	.83	c1	c1, c1	.64	10	6	c1	ci13	1.26	.65	c1	c1, c1	.73	9	10	c1
cr9	1.37	.68	c1	c1, c1	.67	12	4	c1	ci7	1.37	.49	c1	c1!, c2	.73	13	6	c1!
cr15	1.89	.94	c2	c2!, c1	.35	13	10	c1	ci10	1.58	.69	c1	c1, c1	.55	20	25	c2!
cr6	1.89	.88	c2	c2!, c3	.36	19	14	c2!	ci11	2.06	1.03	c2	c2!, c1	.32	16	12	c1
cr2	2.16	.96	c2	c2!, c3	.34	31	52	c3	ci15	2.21	.86	c2	c2, c2	.37	23	24	c2
cr5	2.21	.79	c2	c2, c2	.40	20	13	c2	ci5	2.21	.86	c2	c2!, c3	.37	21	15	c2!
cr11	2.26	.73	c2	c2, c2	.42	19	10	c2	ci1	2.53	.70	c3	c3!, c2	.39	19	12	c2
cr4	2.32	1.00	c2	c2!, c3	.32	27	26	c3	ci16	2.58	1.02	c3	c3!, c2	.30	23	26	c3!
cr12	2.68	.75	c3	c3!, c2	.40	23	12	c3!	ci14	2.95	.97	c3	c3!, c4	.33	32	38	c3!
cr10	2.84	.60	c3	c3!, c2	.51	25	11	c3!	ci9	3.00	1.05	c3	c3!, c4	.31	134	180	c4
cr16	2.89	.88	c3	c3, c3	.36	21	12	c2!	ci2	3.32	.67	c3	c3, c3	.39	37	36	c4!
cr13	3.53	.84	c4	c4, c4	.59	141	368	c4	ci4	3.32	.75	c3	c3!, c4	.37	36	33	c3!
cr7	3.63	.68	c4	c4, c4	.67	56	63	c4	ci12	3.42	.77	c4	c4!, c3	.55	36	32	c3
cr1	3.68	.58	c4	c4, c4	.73	128	281	c4	ci3	3.42	.77	c4	c4, c4	.55	45	42	c4
cr3	3.84	.37	c4	c4, c4	.92	44	36	c4	ci6	3.53	.77	c4	c4!, c3	.60	43	45	c4!

Id = identification for the task, Rm = rank based on the mean ranking-score of the card sort, Rf = rank based on conceptual frame of reference, rm = rank based on the mean ranking score of the estimated students' time on task (min), ! = deviation from conceptual frame of reference. For card sort: c1 = very simple task [1, 1.66], c2 = simple task (1.66, 2.5], c3 = complex task (2.5, 3.33], c4 = very complex task (3.33, 4]. P (c = ci); confidence 'ci' is correct, ci is based on the mean ranking score of the card sort.

The correspondence with the conceptual frame of reference was 87.5% for very simple tasks, 62.5% for very complex tasks, but only 37.5% for simple tasks and 25% for complex tasks. This pattern was observed for both the Criminal Law and the Civil Law tasks. The tasks, based on their confidence scores (P(c = ci)), could not be clearly attributed to one complexity class; only Criminal Law task 3 (cr3) could be attributed with high confidence ($p < .1$) in one category, namely very complex. Here too there was more consensus for tasks on the extremes of the scale (belonging to either very simple or very complex) than for the two middle categories (either simple or complex).

Estimated students' time on task

Descriptive statistics on students' time on task show participants' ratings for the separate tasks again differing greatly (Table 3.3). Spearman's correlation between the ranks from the card sort and the ranks from estimated students' time on task was .785 ($p < .01$) for the Criminal Law tasks and .921 ($p < .01$) for the Civil Law tasks (Table 3.3). Thus, the complexity rankings resulting from estimated students' time on task and card sort were highly correlated.

Table 3.3
Ratings for criminal and civil law learning tasks (16-points-ranking scale)

Criminal law tasks			Civil law tasks		
Id	RCS	RTm	Id	RCS	RTm
cr14	1	4	ci8	1	2
cr9	3	2	ci13	2	1
cr8	2	1	ci7	3	3
cr15	5	3	ci10	4	6
cr11	7	6	ci11	5	4
cr6	4	5	ci16	8	9
cr2	6	12	ci15	7	8
cr4	9	11	ci1	9	5
cr5	8	7	ci5	6	7
cr12	10	9	ci14	11	10
cr16	11	8	ci4	12	11
cr13	13	16	ci9	10	16
cr10	12	10	ci2	14	13
cr7	14	14	ci12	13	12
cr3	16	13	ci6	16	14
cr1	15	15	ci3	15	15

RCS = rank card sort derived from mean in Kendall's *W* test. RTm = rank time on task derived from mean time on task

Rating criteria

From the results of participants' scores on the 18 assertions about criteria they used for rating the complexity of the learning tasks their – collective - conceptual frame of reference for rating could be derived. Means for those criteria on the 4-point scale ranged from 2.21 (SD = .86) to 3.68 (SD = .48).

From these data, the three most important criteria for their ratings were: (a) quantity of information searched for and combined (M = 3.68, SD = .48), (b) quantity of information given and combined (M = 3.47, SD = .52), and (c) kind of intellectual operations required (M = 3.42, SD = .61).

Participants' time on task

It took the raters approximately five minutes to evaluate each task, including the time needed to read the task (for 16 tasks: M = 73.3 min, SD = 12.5 min). The time needed to estimate 'students' time on task' was about one minute for each task (for 16 tasks: M = 16.7 min, SD = 2.6 min). Here the task was conducted after the ranking-task so the reading time of the task was not taken into account.

Ease of use

Cognitive load on the 9-point categorical scale (1 = very, very easy, 5 = not easy, not difficult, 9 = very, very difficult) can be used as an indication for the ease of using the instruments. Perceived cognitive load values were collected for the card-sorting task (M = 5.68, SD = 1.77), the ranking task (M = 4.47, SD = 1.65) and the 'estimated students' time on task'-task (M = 6.32, SD = 1.89). Cognitive load values showed that students' time on task estimations cause the highest load, but comparing the mean cognitive load values in an independent samples t-test showed that this task was not significantly more mentally demanding than the other tasks (card-sort and ranking). All tasks were low to moderately mentally demanding for participants.

Discussion

Results show that neither specific expertise nor type of university influenced the ratings. These findings show that experts from both sub domains, from different types of universities, and tasks from both sub domains could be used in the following experiments.

Since the participants did not regard their teaching expertise to be significantly more important than their subject matter expertise, one might ask whether raters should necessarily be teachers. If, for example, graduate level Law students make similar ratings to Law teachers, this would allow raters to be more easily recruited (i.e., there are more graduate students than professors and they possibly have more time available) and would make the rating process less costly (i.e., students demand a lower wage than their professors). The effects of teaching expertise versus subject matter expertise were studied in Experiment 3.

Participants' rankings for estimated students' time on task showed results similar to their rankings from the card sort, a result that is consistent with the findings by other researchers (Maynard & Hakel, 1997; Winne, 1997). In our view, time-estimations can only be used to determine relative task complexity and not absolute task complexity since tasks in the field of Law often require reading large amounts of information. In other words, time on task estimations can only be used to predict the complexity of a learning task provided one avoids a rating artifact as 'length of task'. This criterion was met in the first experiment.

The task-complexity ratings showed too much variance for confident rating. Separate tasks in the intermediate categories had the lowest confidence values; the extremes showed more consensus. Participants may need more explanation on these categories for more uniform ratings. Providing anchor tasks might improve the consensus in their ratings of the separate tasks. Experiment 2 was designed to determine such anchor tasks.

The highly significant value of Kendall's coefficient (W) in the card-sort task shows that the participants sorted the items on the basis of the same criteria. Based upon the results obtained on rating criteria, the conceptual frame of reference used by the participants appears to coincide with the conceptual frame of reference used in this study, which primarily centers on the criterion 'kind of intellectual operations required'. Participants indicated this among their three most important criteria for their ratings. Their ratings were quite often completely in line with the conceptual frame of reference and never showed a deviation of more than one class.

The results on time needed for rating and ease of instrument use show that the participants were able to make their ratings quickly and do not find this to be mentally demanding. This is very encouraging since the goal of this developmental study is to develop an instrument that is quick and easy to use.

Experiment 2

Experiment 2 was carried out to determine anchor tasks that could be used in Experiment 3. Such anchor tasks to benchmark a complexity category are expected to positively support the rating process in making it more reliable. Since the second experiment was conducted in a fashion largely similar to the first, the methodology will be treated in less detail.

Method

Participants

Twelve Law teachers (8 Criminal Law, 4 Civil Law) employed at Dutch universities returned their results in this experiment (response rate 80%). Nine of them also participated in Experiment 1.

Materials

The materials now consisted of descriptions of 24 Law learning-tasks (20 Criminal Law, 4 Civil Law; none of which was used in the previous experiment) and a series of questionnaires similar to Experiment 1. The tasks provided had - according to the development team - an equal distribution within the conceptual frame of reference (i.e., six tasks for each category). The questionnaire for gathering data on the participants' conceptual frame of reference for rating was adapted for this experiment. Nineteen assertions (18 from Experiment 1 and 1 new assertion) were now scored on a 6-point categorical scale to allow for more sensitive analyses. In this second experiment participants had to sort 24 tasks instead of 16.

Design and procedure

The design and procedure was the same as for the first experiment.

Results

Card sort and anchor tasks

Kendall's W was calculated for both the 24-point ranking scale ($W = .821$, $p < .01$) and the 4-point ranking scale ($W = .646$, $p < .01$).

The rating results for the separate law learning tasks to choose anchor tasks are presented in Table 3.4. The right half of this Table presents results from Experiment 3 and can be ignored at this moment. The order of the tasks, from very simple to very complex, was based upon the mean rating scores. The classification of a task in one of the four categories on the basis of the mean score or on the basis of the median were, except for task 24, the same for all tasks. The data showed that the ratings for the separate tasks differed quite a lot making it impossible to select more than one anchor task per intermediate category.

Tasks 10 and 1 could be attributed with high confidence ($p > .95$) to category 1 (very simple). Task 10 was chosen as anchor task for this category since task 1 might be too obvious as representative for this category. Both tasks 16 and 2 could be attributed with high confidence to category 4 (very complex) ($p > .95$). Task 2 was chosen as anchor task for this category. For the intermediate categories, the task with both the highest probability of correctly belonging to a category (P(c = ci)) and where the rank based upon the mean and the conceptual frame of reference (Rm = Rf), was used as anchor task. For category 2 this was task 5 and for category 3 task 17. It was not possible to choose anchor tasks for those categories with a confidence level of 90% or higher.

Table 3.4
Ratings for law learning tasks in Experiments 2 and 3 (4-points ranking scale)

Law tasks (Experiment 2)						(same tasks in Experiment 3) Teachers						Students					
Id	*M*	*SD*	*Mdn*	Rm, Rf	P(c = ci)	*M*	*SD*	*Mdn*	Rm, Rf	P(c = ci)	rank	*M*	*SD*	*Mdn*	Rm, Rf	P(c = ci)	rank
t1	1	0	c1	c1, c1	1	1	0	c1	c1, c1	1	1	1	0	c1	c1, c1	1	1
t10	1.08	.29	c1	c1, c1	.98	anchor task											
t9	1.17	.39	c1	c1, c1	.90	1.31	.75	c1	c1, c1	.68	3	1.17	.39	c1	c1, c1	.91	2
t12	1.42	.51	c1	c1, c1	.68	1.15	.55	c1	c1, c1	.82	2	1.17	.39	c1	c1, c1	.91	3
t20	1.42	.51	c1	c1, c1	.68	1.46	.66	c1	c1, c1	.62	4	1.33	.65	c1	c1, c1	.70	4
t8	1.67	.65	c2	c2, c2	.41	1.62	.65	c1	c1!, c2	.52	5	1.33	.65	c1	c1!, c2	.70	5
t18	1.92	.67	c2	c2!, c3	.30	1.92	.64	c2	c2!, c3	.48	6	2.17	.72	c2	c2!, c3	.44	6
t23	2.08	1.08	c2	c2!, c1	.30	2.54	1.05	c2*	c3!, c1	.29	10	2.25	.97	c2	c2!, c1	.33	7
t22	2.17	1.03	c2	c2, c2	.32	2.54	1.13	c3	c3!, c2	.27	11	2.42	.79	c2	c2, c2	.37	10
t21	2.25	.75	c2	c2!, c3	.41	2.62	.65	c3	c3, c3	.44	13	2.75	.87	c3	c3, c3	.36	12
t5	2.33	.78	c2	c2, c2	.39	anchor task											
t11	2.42	.79	c2	c2, c2	.37	2.08	.64	c2	c2, c2	.49	7	2.25	.97	c2	c2, c2	.33	8
t14	2.50	.90	c2	c2!, c3	.32	2.15	.80	c2	c2!, c3	.40	8	2.33	.65	c2	c2, c2	.45	9
t15	2.67	.49	c3	c3!, c2	.55	2.46	.78	c2	c2, c2	.37	9	2.50	.67	c2	c2, c2	.39	11
t6	2.67	.78	c3	c3, c3	.39	2.77	.44	c3	c3, c3	.50	14	2.83	.72	c3	c3, c3	.43	13
t17	3.00	.60	c3	c3, c3	.51	anchor task											
t19	3.00	1.13	c3	c3!, c4	.28	3.54	.88	c4	c4, c4	.59	16	3.42	.67	c4	c4, c4	.55	16
t7	3.17	.72	c3	c3!, c2	.41	2.62	.87	c3	c3!, c2	.35	12	2.92	.79	c3	c3!, c2	.40	14
t24	3.41	.51	c3*	c4, c4	.56	3.92	.28	c4	c4, c4	.98	20	3.50	.80	c4	c4, c4	.58	17
t13	3.5	.90	c4	c4, c4	.58	3.69	.63	c4	c4, c4	.72	18	3.75	.62	c4	c4, c4	.75	18
t3	3.58	.51	c4	c4!, c3	.69	3	.82	c3	c3, c3	.38	15	3.17	.58	c3	c3, c3	.49	15
t4	3.75	.45	c4	c4, c4	.82	3.54	.52	c4	c4, c4	.66	17	3.92	.29	c4	c4, c4	.98	20
t16	3.92	.29	c4	c4, c4	.98	3.77	.44	c4	c4, c4	.84	19	3.83	.39	c4	c4, c4	.90	19
t2	3.92	.29	c4	c4, c4	.98	anchor task											

Id = identification for the task, Rm = rank based on the mean ranking-score of the card-sort, Rf = rank based on conceptual frame of reference, ! = difference with conceptual frame of reference, * = difference between classifications based on mean or median. c1 = very simple task [1, 1.66], c2 = simple task (1.66, 2.5], c3 = complex task (2.5, 3.33], c4 = very complex task (3.33, 4]. P(c = ci); confidence 'ci' is correct, ci is based on the mean ranking score of the card sort.

Estimated students' time on task

Spearman's correlation between the ranks from the card sort and the ranks from estimated students' time on task was .95 ($p < .01$). Again, the complexity rankings resulting from estimated students' time on task and the card sort were very similar.

Rating criteria

The means on the 6-point scale of the participants' scores on 19 assertions about criteria for rating the complexity of the learning tasks ranged from 3.08 ($SD = 1.31$) to 5.33 ($SD = .65$). The three most important criteria for their ratings were: (a) kind of intellectual operations required ($M = 5.33$, $SD = .65$), (b) quantity of information searched for and combined ($M = 5.08$, $SD = .67$), and (c) quantity of juridical judgment ($M = 4.92$, $SD = .79$).

Participants' time on task

The 'card sort and ranking'-task including reading time, took approximately four minutes per Law task (for 24 tasks: $M = 104.1$ min, $SD = 51.4$ min). Estimating 'students' time on task' took less than one minute for each Law task (for 24 tasks: $M = 15.5$ min, $SD = 7.2$ min). As was the case in Experiment 1, this task was conducted after the ranking-task so the reading time was not taken into account.

Ease of use

Perceived cognitive load values for the card-sorting task (M = 6.08, SD = 1.31), the ranking task (M = 5.67, SD = 1.67) and the 'estimated students' time on task'-task (M = 6.00, SD = 1.65) were collected. Comparing those mean cognitive load values in an independent samples t-test showed that all tasks were of comparable load and imposed a moderate mental demand on the participants.

Discussion

The most important result was that all of the findings from Experiment 1 were replicated in Experiment 2.

Despite the highly significant value of Kendall's W, the results in this experiment again show that participants differed considerably in their ratings for the separate tasks, especially for the tasks belonging to the intermediate categories. As shown in Experiment 1, these differences cannot be attributed to participants' expertise.

The criteria indicated as most important - as was the case in Experiment 1 - closely resemble those that the development team used for constructing the conceptual frame of reference. Nevertheless, the differences of participants' scores within this conceptual frame of reference become especially manifest for categories 2 and 3. This is most probably the result of the earlier discussed breadth of intellectual operations within these two classes, which makes it difficult to unequivocally attribute a task to a certain complexity. Another complication is that the relative contribution or weight of *each* of the four discriminated intellectual operation classes to complexity is unknown. Using anchor tasks to benchmark a category could help the classification.

Experiment 2 was needed to choose anchor tasks for each category. For categories 1 and 4 this choice could be made confidently. For the two intermediate categories (2 and 3) less unequivocal anchor tasks had to be chosen.

Experiment 3

Experiment 3 had three goals. The first goal was to determine the effect of providing anchor tasks for rating Law tasks. To this end, the results of the teacher group of Experiment 2 (without anchor tasks) were compared with the teacher group of this experiment (with anchor tasks). It was expected that using anchor tasks would result in more agreement about the ratings of the separate tasks. The experiment was also conducted to determine whether additional (better) anchor tasks could be identified. Finally, it investigated whether Law teachers and graduate level Law students would rate sophomore Law learning tasks in a similar manner. It was hypothesized that they would make similar ratings since all participants had encountered the to-be-rated tasks during their own study, and that the lack of experience and pedagogic insight of the graduate students would be compensated by the fact that they only recently covered the material.

Method

Participants

Two groups of teachers (Group 1, 2) and one group of graduate students (Group 3) were involved in Experiment 3. The 12 Law teachers (8 Criminal Law, 4 Civil Law) from Experiment 2 constituted Group 1 and functioned as a control group here. Although this group is reported upon in Experiment 3 as if it was a part of the experiment, it was not. The results were obtained in Experiment 2. Since all instruments and procedures were equivalent in Experiment 2 and Experiment 3, this was considered methodologically valid. Group 2 consisted of an additional 13 law teachers (7 Criminal Law, 6 Civil Law) employed at Dutch universities, six of them also participated in Experiment 1. Group 3 consisted of 12 graduate level Law students who had not yet chosen a specialty. Groups 2 and 3 were used to measure the effect of the anchor tasks on the rating results, with Group 1 (from Experiment 2) functioning as control group (no anchor tasks).

Materials

As four anchor tasks were chosen out of 24 tasks in Experiment 2, the remaining 20 tasks were used. Groups 2 and 3 received materials identical to the previous experiment with the exception of a separate package with four anchor tasks clearly labeled with the category of complexity they represented. The participants were also questioned on the perceived usefulness of the anchor tasks for rating.

Design and procedure

The procedure was, with exception of the anchor tasks, the same as for the preceding experiments except that the participants from Groups 2 and 3 were instructed to use the anchor tasks when carrying out the 'card-sort' and 'task-ranking' tasks.

Results

The results for participants' time on tasks and ease of use concurred with the earlier experiments and will not be separately reported here.

Anchor tasks

The card sort rankings constructed on the basis of the mean ranking scores for the separate tasks for both groups of teachers (Group 1, 2) were compared in a non-parametric correlation test (Spearman's correlation = .919, $p < .01$). Both groups of teachers made similar card sort rankings for the learning tasks included in both experiments. The anchor tasks (right half of Table 3.4) affected neither the variance in rating results nor the confidence in the ranking.

Card sort

Kendall's W was calculated for both the 20-point ranking scale ($W = .693$ (teachers), $W = .791$ (students)) and the 4-point ranking scale ($W = .628$ (teachers), $W = .671$ (students)). All coefficients are significant at the 1% level of probability.

Table 3.4 shows the rating results for the separate tasks by both groups of teachers. The left half presents the scores of the teachers in Experiment 2; the results for the teachers and students in Experiment 3 are presented in the right half. The consensus among teacher-participants for very simple and very complex tasks was - as was the case in the previous experiments - larger than for the two intermediate categories. The consensus was not improved by the anchor tasks. It was still not possible to choose anchor tasks for the intermediate categories with a confidence level of 90% or higher. The correspondence with the conceptual frame of reference was 80% for tasks in category 1, 100% for tasks in category 4, but only 40% for tasks in category 2 and 60% for tasks in category 3.

Participants' experience

Spearman's correlation coefficient between the ranks from the card sort for Law teachers and graduate level Law students was .968 ($p < .01$), showing that both groups rated the tasks very similarly.

Estimated students' time on task

Spearman's correlation between the ranks from the card sort and the ranks from estimated students' time on task was .904 ($p < .01$) for teachers (Group 2) and .961 ($p < .01$) for students. Thus, complexity rankings resulting from estimated students' time on task and the card sort were for both groups very similar.

Rating criteria

The means for the teachers' (Group 2) on the 19 assertions on rating criteria ranged from 2.85 (SD = 1.07) to 4.92 (SD = .64). The students' means ranged from 3.00 (SD = .85) to 5.17 (SD = .83).

The three most important criteria for the teachers' were: (a) kind of intellectual operations required (M = 4.92, SD = .64), (b) sub domain of law (M = 4.92, SD = .86), and (c) quantity of information searched for and combined (M = 4.92, SD = 1.04). Students used the criteria: (a) quantity of information searched for and combined (M = .17, SD = .83), (b) kind of intellectual operations required (M = 5.08, SD = .67), and (c) own experience as a student with these tasks (M = 5.00, SD = .95). Both groups emphasize the criterion around which the conceptual frame of reference is primarily centered (i.e., intellectual operations). There were no significant differences in their scores on 18 of the 19 assertions. A t-test for independent samples showed a significant difference between the groups for the means of 'own experience as a student with these tasks' ($t(23) = -3{,}18$, $p < .01$); graduate level students rated this criterion significantly higher than teachers.

Participants' opinion on the usefulness of anchor tasks

Statistical characteristics for participants' scores on the 6-point scale on the usefulness of the anchor tasks for their ratings are presented in Table 3.5. Participants regarded the anchor tasks to be of only limited use.

Table 3.5
Usefulness of anchor tasks for rating

	teachers			students		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*
a1= anchor task for category 1 (very simple)	2.80	1.40	3	4.08	1.31	5
a2= anchor task for category 2 (simple)	2.55	1.37	2	3.25	1.06	3
a3= anchor task for category 3 (complex)	2.64	1.50	2	3.25	.87	3
a4= anchor task for category 4 (very complex)	2.46	1.29	2	3.58	1.38	4

Participants indicated on a 6-point categorical scale their agreement (1 = totally disagree, 6 = totally agree) with the assertion "anchor task[i] was useful for the classification of the tasks in category [i]"([i] = 1, 2, 3 or 4).

Discussion

Experiment 3 investigated whether the use of anchor tasks would result in more confident ratings for the individual tasks. This appeared not to be the case for the teachers. It is not certain whether this also holds for graduate level students, as students' rating results without the use of anchor tasks were not collected. Again, participants differed in their ratings for the separate tasks, especially for the tasks belonging to the intermediate categories.

Experiment 3 also did not result in discovering additional or better anchor tasks. As it was not feasible to unequivocally identify representative anchor tasks for the intermediate categories in Experiment 2, one could not expect a panacea from using them in Experiment 3.

Raters thought *all* anchor tasks to be of not much help in classifying the learning tasks. This is supported by the data that show that the classifications of the tasks for *all* categories did not have less variance when anchor tasks were available.

A positive result is the finding that graduate level students can apparently be used for rating learning tasks. Students rank the tasks in a similar way and report using a similar conceptual frame of reference for rating, which was also in line with the conceptual frame of reference as was used by the researchers. This simplifies the work of the educational developer for determining the complexity of learning tasks in MmPs since the pool of students is larger than the pool of teachers and the fees that need to be paid are lower. Teaching experience does not result in different or more valid ratings. The most plausible explanation for this finding is that graduate students' more recent experience with the tasks and content compensates for a lack of teaching experience and/or a lesser depth of knowledge. This is also in line with the finding that students' report that their own experience with these tasks is one of the three most important rating criteria. Teachers report this criterion to be significantly less important for their ratings.

The results of this experiment reconfirm the results from earlier experiments on card sort rankings and estimated student-time on task rankings.

General discussion

The most important conclusion to be drawn from this study is that the instrument developed is reliable, easy to use, and does not require any specific training. Results from all three experiments show that raters apply the same standard in ranking learning tasks, using criteria similar to those in the conceptual frame of reference. The most relevant expertise for rating the learning tasks is the raters' own experience as a student with the tasks. Ratings were not influenced by the area of expertise or type of university (Experiment 1). Graduate level Law students and Law teachers also rated the tasks similarly (Experiment 3). Complexity rankings based on estimated students' time on task and card sorting are similar. Both methods of ranking are relatively fast and easy to carry out. Provided that length of task is controlled for estimating students' time on task is as reliable as, but faster than card sorting.

Raters' classification of learning tasks belonging to the intermediate categories were not always in agreement with the conceptual frame of reference. Compared to the extremes, the raters were also less confident about their ratings. If deviations occur between raters' classifications and the conceptual frame of reference, the deviations are mostly for tasks in the intermediate categories and seldom for the extremes. The deviation is never larger than one category.

There are four possible explanations for these observations. First, the size of a task (amount of cognitive operations) versus the sort of cognitive operations (level) makes it impossible to operationalize complexity in the same way for categories 2 and 3. As a result, and contrary to the assumptions, categories 2 and 3 are not disjunctive, but partially overlap. Second, the tasks included in the experiments were not unmistakably representative for the conceptual frame of reference because they were taken from existing instructional materials. Therefore, it cannot be sure that the anchor tasks were fully representative for their class. This probably had the largest effect on categories 2 and 3 because the attribution of intermediate categories of a scale is always more difficult than the attribution of categories at both ends (P.G. Swanborn, personal communication, December 4, 2001), especially if the complexity of learning tasks has a normal distribution. Third, the conceptual frame of reference was quite abstract and therefore the anchor tasks might have been too limited (the conceptual frame of reference and the function of the anchor tasks were not made explicit to the raters). Again, this may have had the largest effect on categories 2 and 3. Finally, since complexity is a multidimensional concept, the one-dimensional approach chosen in this study concentrating only on intellectual operations might explain these small classification anomalies. Below, the four possible explanations will be treated in more detail.

The first reason for classification problems is that the way complexity is made operational for the categories 2 and 3 is contaminated by the size of a task (i.e., the number of cognitive operations that need to be carried out) versus the type of cognitive operations (i.e., the level of the cognitive operations to be carried out). As a result, complexity categories 2 and 3 do not perfectly match cognitive operation levels 2 and 3. The classes in the conceptual frame of reference for intellectual operations are expected to be disjunctive and therefore not to show overlap. Although Merrill's theory (1987) was used, several other models for classifying cognitive operations show the same four categories (Anderson & Krathwohl, 2001; De Block, 1975; Crombag et al., 1979; Lewy & Báthory, 1994). Nevertheless, these categories do not specifically represent levels of complexity. It is known from other models for determining task complexity that it is impossible to objectively weigh identified task characteristics (Campbell, 1988; Wood, 1986). This may also be the case for the categories in the conceptual frame of reference. The categories can be clearly identified, but they are not completely in line with increasing complexity. This is especially true for tasks belonging to category 2 ("understand a generality") and category 3 ("use"). In line with this is that the size of a task within these two categories of intellectual operations can vary enormously. Thus, in some circumstances understanding a generality (e.g., understanding the concept Justice) can be more difficult than using a particular piece of knowledge (e.g., applying a simple procedure for finding the maximum punishment that can be applied for a certain crime). In other words, a large task requiring low level cognitive operations might be more complex than a small task requiring higher level cognitive operations.

The second reason for classification problems, especially for category 2 and 3 tasks, is that the to-be-rated tasks were not specifically developed for the experiments and were therefore not unmistakably representative for the conceptual frame of reference. In addition, this may explain the marginal effect of the use of anchor tasks on the variance of the rating results, since the anchor tasks, *ipso facto*, also were not specifically developed as a representative task for a particular category. Therefore, it cannot be ruled out that they were not unequivocally representative of one category, especially since participants in Experiment 2 did not fully agree on the anchor tasks chosen for Experiment 3. In other words, the anchor tasks were not always perceived of as good representatives for their categories. Furthermore, there were no explanations given as to why particular anchor tasks represented a particular category. Together, these considerations may explain the limited value of the anchor tasks.

The third explanation for the classification problems is that it was chosen to use *only one* anchor task to represent a category. The use of only one anchor task might have resulted in an overly limited category representation. Indeed, participants indicated for each category that the anchor task representing this category was not of much help to them during their classifications. The idea that anchor tasks could have been too limited is also supported by the data of participants' rating criteria. Despite a close resemblance of the participants' criteria for rating the tasks with the conceptual frame of reference in this study, the raters nevertheless show a high variance in the ratings for the separate tasks. Since the conceptual frame of reference was not made explicit to the participants, the function of the anchor tasks might also have been too veiled for them and therefore too abstract. Anchor tasks showing the upper and lower limits of a category as well as explication of why the task is considered to be representative for a category (i.e., the conceptual framework) might have been more helpful.

The fourth reason for classification problems might be that it was chosen to operationalize complexity as a one-dimensional concept, namely on the dimension *intellectual operations*. As is the case for intelligence, complexity can also be regarded as a multidimensional concept. Guilford (1982) proposed a factor-analytic model of intelligence consisting of 150 independent abilities that result from the interaction of five types of contents, five types of operations, and six types of products. Sternberg (1985) went "Beyond IQ" offering a "triarchic theory of human intelligence" with three components: analytic (academic) intelligence, creative intelligence, and practical intelligence. Yet the fact that it is commonly accepted that intelligence is a multidimensional concept, this does not preclude research on any one of those dimensions. Complexity too is probably a multidimensional concept, with other dimensions being *quantity of information* searched for and combined (more quantity is more complex), *field or discipline* (some disciplines are more complex than others), *symbolic system of task-formulation* (text and/or graphics, animations; some 'languages' and/or symbol systems are more complex than others), *preferences and styles of the receiver* (some people are text oriented, other iconic), et cetera. Since intellectual operations appear to be hierarchical, causing their relative contribution to a multidimensional complexity construct to increase in the higher categories, we expect intellectual operations to be an important dimension in determining objective task complexity. In this, we lean on the work of important theorists in the field such as Bloom (see, e.g., Anderson & Krathwohl, 2001) and Merrril (see, e.g., Merrill, 1987). However, more research is needed in which a multidimensional theoretical construct on learning task complexity should be tested.

Two promising lines for further research can be distinguished for measuring the complexity of learning tasks. First, research should be conducted as to whether the explanation of the conceptual frame of reference together with an explanation of the way that anchor tasks fit within this frame, would result in more confident ratings for the separate tasks. In the present study, the conceptual frame of reference was not made explicit, because it was yet unknown if participants' ratings would be similar to this frame. It is now known that participants use similar criteria. Explaining the conceptual frame of reference in future work could result in more consensus between ratings for the separate tasks.

A second line includes implementing the current instrument for measuring task complexity in an Instructional Design model to find out whether this results in improved, more effective designs and products. Chapter 2 describes an ID-model whose application reduces the complexity of learning tasks in competency-based learning environments through a multiple-step whole-task approach, while not sacrificing the authenticity of the learning experience. Figure 2.2 in Chapter 2 presents this two-phase six-step ID-model. As described in detail in Chapter 2, Phase 1 deals with Cognitive Task Analysis (steps 1 through 3) and Phase 2 with actual training design (steps 4 through 6), resulting in a detailed blueprint for the learning environment. The six steps are: (1) decomposing the complex skill, (2) determining task complexity, (3) identifying Systematic Approaches to Problem solving (SAPs), (4) sequencing problems on the micro-level, (5) choosing problem formats, and (6) choosing the number of phases of SAPs that will be presented to learners as process worksheets. Steps 2 and 6 involve the determination of task complexity. Up till now, teachers had to use a subjective, intuitive measure of task complexity because there was no available instrument for objectively measuring task complexity. Now that this instrument has been developed, experiments will be conducted to determine the optimal number of phases within process worksheets.

From a practical point of view, it is important to develop instruments for the measurement of task complexity in other domains. The conceptual frame of reference presented in this chapter offers a good starting point because the identified intellectual operations are not exclusive to the domain of Law. It should be stressed once again that an effective use of such an instrument presupposes that students confronted with the rated learning tasks have roughly the same prior knowledge. Indeed, an instrument for measuring task complexity has only limited value if students differ greatly in prior knowledge. A current trend in education is to develop personalized, student-centered instruction that takes differences in prior knowledge into account. Then, at first sight, it seems to be of little use to develop instruments that measure task complexity in advance, that is, before the learning tasks are actually presented to the learners. Collecting data on subjective task complexity from learners seems a workable solution for tailoring the instructional material "on the fly".

Research on instruments for measuring learning task complexity will become increasingly important because the demand for competency-based learning environments and MmPs is still growing. In such environments (e.g., Wöretshofer et al., 2000), it is of utmost importance to carefully adjust the complexity of learning tasks and the number of phases within SAPs that support learners in performing those tasks to the target learners. Measurement instruments for task complexity support instructional designers in this process, yielding better support for learners and more effective learning.

Chapter 4 - Optimizing the number of phases in learning tasks for complex skills*

Whole tasks are often too difficult for novice learners for learning complex skills. The common solution is to split up the problem solving process of the learning tasks into phases. The number of phases must, in turn, be optimized for efficient and effective learning. It is hypothesized that students solving a whole task with an intermediate number of phases will outperform students exposed to either no phases at all (i.e., they learn to solve the whole task in only *one* phase) or a high number of phases. Sophomore Law students (N = 35) were randomly assigned to three computer-delivered versions of a multimedia training program for learning to carry out a plea in court. As hypothesized, an intermediate number of phases constitute most effective training showing that the number of phases can indeed be optimized for learning.

Introduction

Learning in Multimedia Practicals takes place in a self-contained electronic learning environment. Such practicals provide context-relevant practice to students for attaining complex skills such as diagnosing diseases, literature searching, modeling stress-factors that cause burn-out, or preparing a plea in court (Brown, Collins, & Duguid, 1989; Nadolski, Kirschner, van Merriënboer, & Hummel, 2001; Westera & Sloep, 1998). These practicals are assumed to be instrumental in allowing learners to develop the cognitive schemas necessary for the performance and transfer of complex skills. Many researchers (e.g., Hannafin, Land, & Oliver, 1999; Jonassen, 1999; Mayer, 1999; Merrill, 2002; Stark, Gruber, Renkl, & Mandl, 1998; van Merriënboer, 1997) agree that transfer-oriented learning can best be achieved through the use of realistic *learning tasks* consisting of a task description, an authentic environment to carry out the task, and cognitive feedback on the quality of the task performance.

This, however, is often easier said than done. Realistic whole tasks are often too difficult for novice learners without some form of simplification. A common solution to preclude this problem is first to conceptually model reality (i.e., simplify it) and second to pedagogically model this model (Achtenhagen, 2001). Pedagogical modeling (i.e., didactic specification, see Resnick, 1976) can be achieved through segmentation of the whole learning task into smaller task assignments and thereby dividing the problem solving process into *phases* which are presented in a process worksheet. In other words, after modeling reality, pedagogical modeling streamlines the problem solving process of the whole learning task as it segments this process into phases. For example, in Figure 1.2 in Chapter 1 the whole task of "preparing & pleading a case in court" has been segmented into seven meaningful task assignments. Then, a *process worksheet* (van Merriënboer, 1997) provides the phases to the learner and guides them through the problem solving process. A process worksheet (not filled out in the figure) refers to the task assignments that guide the learners through the distinct phases; it provides a Systematic Approach to Problem Solving (SAP) for the whole learning task. The two-part process is similar to what Achtenhagen (2001, p.364) calls "modeling a model of reality under a didactic perspective". Most tasks can be segmented into smaller segments,

* Based on: Nadolski, R. J., Kirschner, P. A., & van Merriënboer, J. J. G.. Optimising the number of steps in learning tasks for complex skills. Manuscript submitted for publication.

which then can be broken down again and again and which can be solved using the phases presented in the process worksheet. The question is, how to determine what the optimal number of phases is: when do you stop? The central question is thus: Can the number of phases be optimized and will this lead to better task performance and higher task efficiency?

Optimizing the number of phases and providing an accompanying process worksheet brings learning tasks within the reach of the learners' capabilities. Tasks divided into too few phases are often too difficult and mentally demanding for learners to carry out, which hampers learning and subsequent transfer. Learners may either not accurately process the necessary information because they are overwhelmed by the difficulty of the task (i.e., cognitive overload) or may revert to surface processing (i.e., superficial, non-meaningful learning) in order to keep their cognitive load within the threshold limit (Craik & Lockhart, 1972; Sternberg & Frensch, 1991). Tasks divided into too many phases may also hamper learning because of their non-coherency caused by redundant information between phases and/or an excess of details. This too will make them too mentally demanding (Mayer & Moreno, 2002). In addition, learners may regard the phases as too specific for the learning task in question, preventing them from constructing the generalizations or abstract schemas necessary for learning transfer. Thus, like many other instructional design problems (see Clark, 1999), determining the number of phases is an instructional design problem that requires a solution through optimization.

The number of phases may directly influence cognitive processing and schema construction, but may also affect the investment learners bring into the training situation (see Bonner, 1994). In this study, investment is operationalized by a combination of measures of motivation, time on task, and mental effort. The more investment, the better the result. However, this is not absolute. First, there is a limit above which a further increase of investment does not yield further increase in performance. Increasing task motivation is not helpful for a learning task that is so difficult that it is almost impossible to solve. More mental effort only leads to a better result if the task performer is not already cognitively overloaded by too much information. Second, above a certain threshold level, longer time on task also does not lead to a better result. Third, task motivation, mental effort and time on task may be irrelevant if a learning task is extremely easy. Finally, interaction effects might also occur where, for instance, higher motivation leads to a shorter time on task with a better result (Bonner, 1994). This means that it is not only needed to study learning results, but also the learner investment in reaching those results when considering optimizing the number of phases. In this study, investment is determined by measuring task motivation (Bonner, 1994; Maynard & Hakel, 1997), subjectively experienced cognitive load (Paas & van Merriënboer, 1994) and time on task (Karweit, 1984).

Finally, several studies have shown the usefulness of not only determining learning effectiveness, but also the learning *efficiency* of instructional design measures (see Admiraal, Wubbels, & Pilot, 1999; Kalyuga, Chandler, & Sweller, 1998; van Merriënboer, Schuurman, de Croock, & Paas, 2002). Although efficiency can be operationalized in many different ways, higher efficiency always indicates either equivalent results at lower investment, higher results at the same investment, or, ideally, higher results at lower investment. When studying the effects of the number of phases, it is thus important to study both the *task results* (i.e., task performance) and the investment associated with reaching those results (based on motivation, cognitive load, and time on task) to determine the *task efficiency*.

This study is designed to examine the effects of the number of phases on the performance and efficiency of learning tasks and transfer tasks. Three conditions (no phases, an intermediate number of phases, and a high number of phases) are compared for Law students learning to prepare a plea in court. The first hypothesis is that an intermediate number of phases will lead to highest *performance* for the *learning tasks*, while no significance differences between the other two conditions will be present. The second hypothesis is that a high number of phases will show lowest *transfer performance*. The final hypothesis is that *task efficiency* will be lowest for a high number of phases and highest for an intermediate number of phases.

Method

Participants

Thirty-five students enrolled in the experiment and were randomly assigned to the three conditions (no phases; *n* = 11, intermediate number of phases; *n* = 12, high number of phases; *n* = 12). All participants (22 female, 13 male; mean age = 22.8 years, *SD* = 3.5) were sophomore Law students studying at Dutch universities. None of the participants had prior plea experience. Comparability of students with respect to domain knowledge was assured since first year law curricula of Dutch universities are practically identical, both with respect to courses and textbooks. Comparability of plea experience was assured by a background questionnaire that will be discussed later.

Materials

Learning materials.

The Multimedia Practical *Preparing a plea* (Wöretshofer et al., 2000) was adapted for this research. The goal of the practical is to prepare students to carry out a plea in court. Students receive multiple-phase whole-task training using a high-variability sequence of learning tasks (one compulsory learning task and two additional, non-compulsory learning tasks). Support is faded as the learners gain expertise, beginning with concrete modeling examples through working with tasks with process worksheets (see Nadolski, Kirschner, van Merriënboer, & Hummel, 2001). The non-compulsory learning tasks enable variability of practice. Three versions of the practical were produced: one with no phases (a whole task = Min-condition), one with four phases (intermediate number of phases with feedback between the phases = Int-condition), and one with nine phases (high number of phases with feedback between the phases = Max-condition). For all versions, objective task complexity of the assignments (1 = very simple, 2 = simple, 3 = complex, 4 = very complex) was determined by 32 different participants using the task-complexity instrument developed by Nadolski, Kirschner, van Merriënboer, and Wöretshofer (in press). The mean objective complexity for the tasks was 3.6 (*SD* = .7) for the Min-condition, 2.9 (*SD* = .4) for the Int-condition, and 2.0 (*SD* = .4) for the Max-condition.

As indicated in Figure 4.1, all versions first presented identical information and support tools. The tools included modeling examples of persons conducting a plea, a "plea checker" to analyze pleas, discussions of ethical issues in pleading a case, numerous tips for communicative aspects in pleading a case, and judicial-procedural aspects of plea preparation. Participants had all necessary documents available while working on the learning tasks. The material for the learning tasks also contained process worksheets that differed for each condition. A specific task assignment was available for each phase included in the process worksheet. The phases in the several versions are related. The task assignment in Phase 1 of the version used for the Intermediate condition, for example, aims at an outcome that is comparable to the outcome for Phase 1 up to including Phase 5 in the Maximum condition. If two phases followed each other, the latter phase always included cognitive feedback on the previous phase (between-phase support) and a new task assignment.

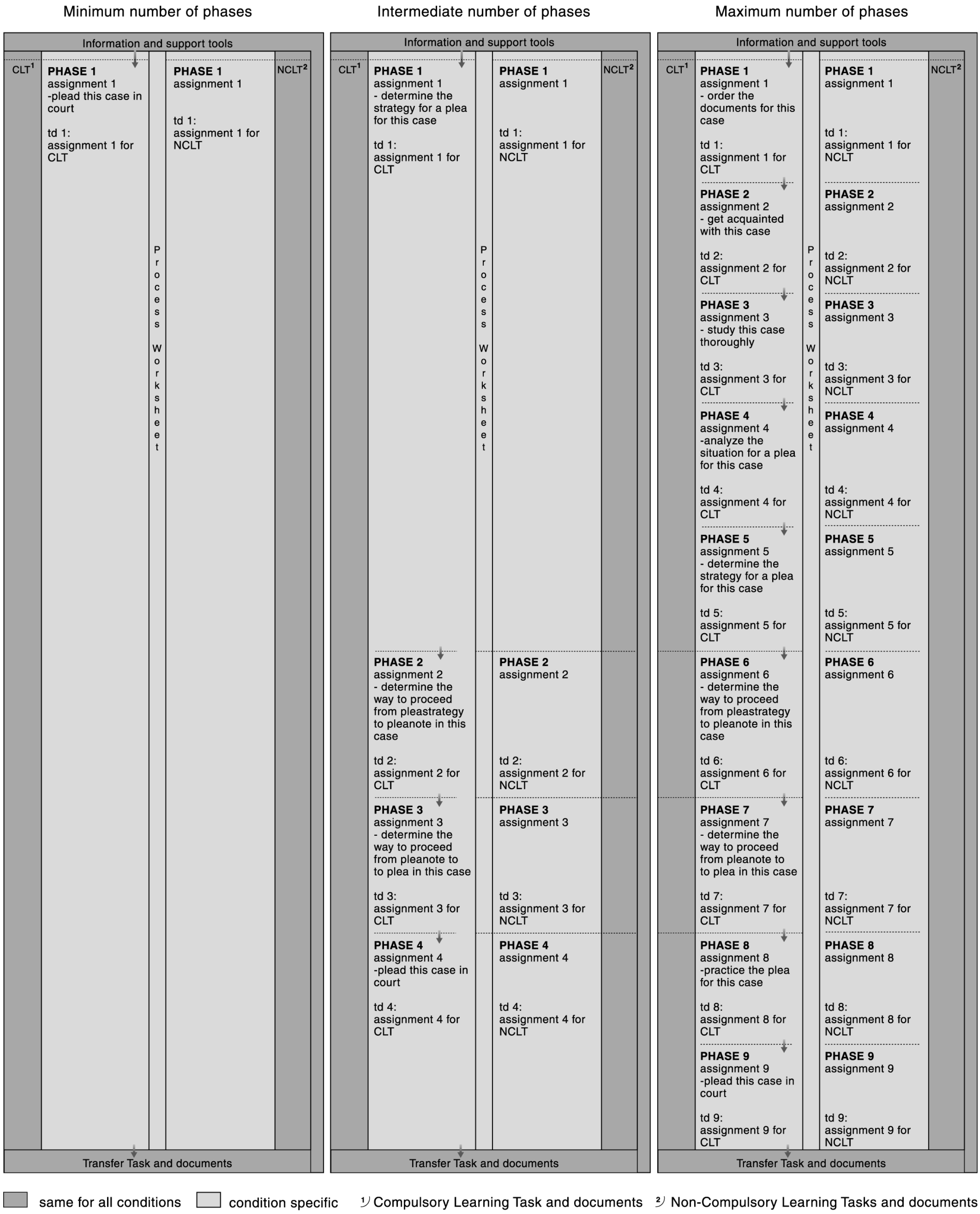


Figure 4.1
Overview of practical material for all conditions.

All phase-information provided remained available for all following phases. After working on the learning task(s), participants received one transfer task. The material for the transfer task was the same in all conditions.

Tests.

Two instruments were developed to measure the coherence, content, and communicability of participants' plea performances on both the compulsory learning task and the transfer task. Elements related to these three performance measures were rated on 4-point scales (0 = poor, 1 = weak, 2 = sufficient, 3 = good). Examples of items for coherence were: 'starts with a conclusion to be drawn from the plea', 'provides a content overview of the plea before going into details', 'indicates how successive arguments relate to relevant facts' and - for each major point - 'provides conclusion with respect to major point [x]'. Since the coherence of the plea is directly related to the number of points needed to be made by the "pleader", the number of items differs for the two instruments. Coherence for learning task and transfer task was scored on 7 respectively 9 items (Cohen's Kappa is .7 and .6 respectively). Content for learning task was scored on 39 items, content related to transfer task was scored on 46 items (Cohen's Kappa is .6 and .7 respectively) and communicability was scored on 14 items for both instruments (Cohen's Kappa is .6 and .7 respectively). Existing plea-measurement instruments (e.g., Edens, Rink, & Smilde, 2000) were regarded too general to be used here.

Also, a series of questionnaires was used to gather data on:

1. Prior plea experience (e.g., prior written and oral presentation skills, membership of debating club), computer literacy, attitude towards learning with computers, age, and gender.
2. Time on task per phase.
3. Mental effort. A 9-point categorical scale developed by Paas, van Merriënboer, and Adam (1994) was used to measure the experienced cognitive load of each of the constituting phases for the learning task (before and after each phase respectively), the learning task itself and the transfer task. Experienced cognitive load was used as an indication for mental effort; the lower the experienced cognitive load, the less mentally demanding the task.
4. Motivation per phase. A 3-item 7-point Likert scale taken from Maynard and Hakel (1997) was used. The items were: "I was motivated to perform well on this task assignment", "This task assignment was interesting to me", and "I put a lot of effort into coming up with the best possible solution".
5. Satisfaction/perceived efficacy. Participants indicated their satisfaction with the quality of the instructional materials by giving their opinion on the adequacy of the number of phases, whether the goal of the practical ("I've learned to conduct a plea") was reached, the relevance of the supportive information, and the relevance of between-phase support.

Finally, as all conditions were computer-delivered, all study behavior was logged and analyzed.

Procedure

Before participating, (potential) participants were informed of the experiment, the Practical (60 study hours in one month) and the needed prior knowledge and skills. Participants were randomly assigned to one of the three experimental conditions and were required to work individually. All learning materials (including instructions) and questionnaires were sent to the participants' home addresses. They were strongly advised to work phase-by-phase since the program offered the possibility of skipping consecutive phases. Logging results showed that the participants did not skip any phases. Within a phase, however, maximum learner control existed. In other words, participants were free to decide if and when to consult phase specific information and how long they would like to spend working on the task assignment in this phase. After two weeks (approximately 30 study hours), participants were required to hold their plea for the compulsory learning task, which was videotaped for later evaluation.

About two weeks later, they were required to hold their plea for the transfer task, which was also videotaped. Since it is legally required for lawyers in the Netherlands to submit a plea note when pleading a case in court (i.e., a memorandum of the oral pleading), participants were required to include a plea note for both tasks. Participants were required to return the questionnaires in a stamped self-addressed envelope and to electronically send their logging results one week after completion of the course. Upon completion of the experiment, participants were thanked for their participation and received the promised remuneration for their participation (circa $ 120). Two judges blindly and independently scored participants' videotaped pleas.

Data analysis and scoring

Two judges blindly and independently scored participants' videotaped pleas. All videotaped pleas were scored by both judges using the performance measurement instruments. All efficiency measures were calculated using the procedure described by Tuovinen and Paas (in press) for determining instructional condition efficiency. They describe a three factor instructional condition efficiency that was extended to a four factor instructional condition efficiency here. In formula: 4 Factor Efficiency = (P-E-T-M) / SQRT (4), where P = performance, E = mental effort, T = time on task, M = motivation. Students' scores on all factors are standardized (the total mean was subtracted from each score and the result was divided by the standard deviation), giving *z*-scores for each factor.

Results

The data collected for determining computer literacy and attitude towards learning with computers showed no differences between conditions. The data collected for prior plea experience confirmed that none of the participants had prior plea experience and thus that the conditions were equivalent. Analysis of the logging results indicated that participants did not make use of the non-compulsory learning tasks.

Performance

Data means and standard deviations for performance on the pleas for the learning task and transfer task are presented in Table 4.1. Analyses of variance showed a significant difference in the *coherence* on the learning task plea between the three conditions ($F(2, 32) = 3.48$, $MSE = .37$, $p < .05$, $\eta^2 = .18$). Contrast analyses, using Bonferroni's correction, revealed a significantly better coherence for the Intermediate condition, as compared to the average coherence for the Minimal and Maximum conditions ($t(32) = 2.5$, $p < .05$ (one-tailed)). No significant difference was found between the Minimal and Maximum conditions ($t(32) = 1.0$, $p = .3$).

Analyses of variance also revealed a significant difference between the conditions with respect to the *content* of the learning task plea ($F(2, 32) = 10.87$, $MSE = .24$, $p < .01$, $\eta^2 = .41$). Contrast analyses, using Bonferroni's correction, revealed a significantly better result for the Intermediate condition, as compared to the Minimal and Maximum conditions ($t(32) = 4.4$, $p < .01$ (one-tailed)). No significant difference was found between the Minimal and Maximum conditions ($t(32) = 1.7$, $p = .1$). There were no differences in the *communicability* of the learning task plea.

Table 4.1
Performance on the learning and transfer tasks

	Min (n = 11)		Int (n = 12)		Max (n = 12)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Learning task						
- coherence	2.09	.8	2.75*	.5	2.33	.5
- content	1.82	.4	2.75**	.5	2.17	.6
- communicability	2.82	.4	2.75	.5	2.5	.5
Transfer task						
- coherence	1.55	.5	1.33	.5	1.33	.5
- content	.91	.5	1.08	.5	.83	.6
- communicability	2.82	.4	2.75	.5	2.5	.5

* p <.05 ** p <.01 for the number of phases.
All performance variables were measured on a 4-point scale (0 = poor, 1 = weak, 2 = sufficient, 3 = good)

Results on the coherency, content and communicability of the plea for the transfer task did not show any significant differences between the conditions.

Time on task

Data means and standard deviations for the time spent on the learning and transfer tasks are presented in Table 4.2. Analyses of variance showed a marginally significant difference in time on task between the three conditions ($F(2, 32) = 2.93$, $MSE = 109757$, $p = .07$, $\eta^2 = .16$) and the predicted pattern of the Max-condition having more time on task than the combined Int- and Min-condition was observed. A contrast test, using Bonferroni's correction, showed that the Max-condition had significantly more time on task for the learning task than the combined Int- and Min-condition, $t(32) = 2.42$, $p < .05$ (one-tailed). There was no significant difference between the Min- and the Int-condition, $t(21) = .18$, $p = .86$.

Table 4.2
Time on task (minutes)

	Min (n = 11)		Int (n = 12)		Max (n = 12)	
	M	*SD*	*M*	*SD*	*M*	*SD*
time on learning task	635	378	660	233	933*	368
time on transfer task	325	104	382	228	473	557

* p < .05, one-tailed contrast between the Max- and combined Int- and Min-condition.
Based on self-report. All spent time was reported in multiples of five minutes. The transfer task was the same for all groups; no support was given.

There were no significant differences on the time on task values for the transfer task between the conditions, although the expected pattern of an increasing time on task with an increasing number of phases, was observed.

Mental effort & motivation

The mental effort values for both the learning task and the transfer task did not show any differences between the conditions (see Table 4.3). In all conditions, both tasks demanded an average mental effort. The motivation values for both the learning task and the transfer task did not differ between the conditions. In all conditions, participants were highly motivated.

Table 4.3
Mental effort and motivation on the learning and transfer tasks

	Min (n = 11)		Int (n = 12)		Max (n = 12)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Learning task						
- mental effort	4.64	1.4	5.5	2.2	5.08	2.4
- motivation	5.79	.9	5.75	.9	6.14	.8
Transfer task						
- mental effort	4.73	1.7	5.92	1.8	5.5	2.5
- motivation	5.33	1.6	5.47	.9	6	.9

Mental effort is measured on a 9-point scale (Paas, van Merriënboer, & Adam, 1994) (1 = very, very low mental effort, 9 = very, very high mental effort). Motivation is measured on a 3-item 7-point scale (Maynard & Hakel, 1997) (1 = very, very low, 7 = very, very high).

Task efficiency

There were no significant correlations between mental effort, motivation, and time on task. Data means and standard deviations for *learning task efficiency* and *transfer task efficiency* are presented in Table 4.4. Analyses of variance on all three measures of quality of the learning task showed a marginally significant difference between the three conditions (learning task coherency: $F(2, 32) = 2.91$, $MSE = 82.9$, $p = .07$, $\eta^2 = .16$; learning task content: $F(2, 32) = 2.91$, $MSE = 84.2$, $p = .07$, $\eta^2 = .53$; learning task communicability: $F(2, 32) = 3.08$, $MSE = 83.3$, $p = .06$, $\eta^2 = .16$). Because explicit hypotheses were formulated, a contrast test, using Bonferroni's correction, was performed which showed that the Max-condition was significantly less efficient on all three measures of quality of the learning task than the combined Int- and Min-condition (learning task coherency: $t(32) = 2.41$, $p < .05$ (one-tailed); learning task content: $t(32) = 2.44$, $p < .05$ (one-tailed); learning task communicability: $t(32) = 2.47$, $p < .05$ (one-tailed)). There were no differences for learning task efficiency between the Int-condition and the Min-condition (learning task coherency: $t(21) = .15$, $p = .89$; learning task content: $t(21) = .10$, $p = .92$; learning task communicability: $t(21) = .27$, $p = .79$). Both conditions were equally efficient.

There were no differences for the conditions with respect to transfer task efficiency.

Table 4.4
Efficiency for the learning and transfer tasks

	Min (n = 11)		Int (n = 12)		Max (n = 12)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Learning task						
- coherence	3.0	10.6	2.4	6.1	-5.2*	10.3
- content	2.9	10.5	2.5	5.9	-5.2*	10.3
- communicability	3.3	10.6	2.3	5.9	-5.3*	10.3
Transfer task						
- coherence	2.8	3.5	.7	6.6	-1.9	15.7
- content	2.2	3.4	.3	6.2	-2.4	15.4
- communicicability	2.3	3.4	.3	6.3	-2.4	15.6

* $p < .05$, one-tailed contrast between the Max- and combined Min- and Int-condition.

Satisfaction / perceived efficacy

Analyses of variance showed a significant difference with respect to participants' opinion of the adequacy of the number of phases, $F(2, 32) = 4.74$, $MSE = .98$, $p < .05$, $\eta^2 = .23$ (see Table 4.5). The Min-condition was most satisfied. Post-hoc tests, using Bonferroni's correction, indicated that the Min- and Max-condition differed significantly ($p < .05$) with respect to participants' opinion of the adequacy of the number of phases. There was a marginally significant difference between the Min and Intermediate-condition ($p = .06$) and there was no significant difference between the Intermediate and Max-condition with respect to participants' opinion of the adequacy of the number of phases. In other words, participants in the Min-condition thought their one-phase approach to be slightly more adequate as those in both other conditions. Participants in both other conditions were not unsatisfied, but indicated that the number of phases could be slightly decreased.

Feelings of efficacy ("I've learned to conduct a plea") did not differ between conditions, which was also the case with respect to their opinions on 'the relevance of the provided supportive information' and their satisfaction as to the between-phases support.

Table 4.5
Satisfaction and perceived efficacy of the training

	Min (n = 11)		Int (n = 12)		Max (n = 12)	
	M	*SD*	*M*	*SD*	*M*	*SD*
number of phases adequacy[a]	3.82*	.9	4.83	.9	5	1.1
"I've learned to conduct a plea"[b]	5.45	1.4	5.67	1.1	5.42	1.1
relevance of supportive info[c]	2.45	.7	2.75	.5	2.66	.8
between-phase-support[d]			2.05	.4	2.30	.6

* $p < .05$ for the number of phases.
[a]The number of phases-adequacy was indicated on a 7-point scale with respect to the assertion "the number of phases is" (1 = far too few, 4 = perfectly all right, 7 = far too many).
[b]The assertion "I've learned to conduct a plea" was scored on a 7-point scale (1 = very strongly disagree, 7 = very strongly agree).
[c]The relevance of supportive info was indicated on a 4-point scale (0 = strongly irrelevant, 3 = strongly relevant).
[d]Satisfaction for between-phase-support was indicated on a 8-items 4-point scale (0 = very dissatisfied, 3 = very satisfied) and could only be collected for the Intermediate and Max-condition.

Discussion

Limiting the number of phases in learning to solve complex tasks leads to optimal learning-effectiveness, as was predicted in hypothesis 1. Too many phases made the learning task less coherent; though time on task was somewhat - although not significantly - increased, no concomitant increase in learning was observed. Too few phases, while as efficient as the optimal condition with respect to amount learnt per unit time, led to a lower performance on the learning task. Since there were no differences with respect to prior knowledge and skills, the results can be attributed solely to the variation in the number of phases. None of the other variables that might affect learning (motivation, increased cognitive load, perceived adequacy, time on task, satisfaction, and so forth) proved to be significant and thus cannot explain this result.

Contrary to the expectations in hypothesis 2, performance on the transfer task did not differ between conditions. As the conditions also did not differ with respect to learner investment in the transfer situation, explanations for the equal transfer-performance require more detailed analyses of the logging data. The data show that learners did not make use of the non-compulsory learning tasks for practicing what they had been taught. This was contrary to the expectation that they would make use of these high-variability practice tasks to practice what the different phases required. Due to this, the learning situation was limited to a single set of learning task assignments. Researchers agree that transfer cannot be expected under such circumstances (e.g., Gagné, Yekovich, & Yekovich, 1993; van Merriënboer, 1997).

In line with hypothesis 3, there was a trend for learning efficiency to be lowest for learners confronted with a high number of phases. While learners in this condition showed a somewhat higher time on task this did not lead to better learning. The observed trend formed the basis for a partial replication of this experiment with only two conditions (i.e., Max and Int) which is reported on in the following chapter (Chapter 5). As the reader will see, this did lead to a main effect. Contrary to the expectations, no differences in time on task were found between learners in the Single-phase and Intermediate number-of-phases conditions. There were also no differences in transfer task efficiency. Explanations for this are similar to those given for the rejection of the transfer hypotheses in hypothesis 2.

The practical implications of this study are quite straightforward. There is clear empirical evidence for the value of optimization the number of phases in learning tasks. Too many phases leads to lower performance and thus does not justify the extra, apparently unnecessary, costs for developing such instructional materials. In other words, development costs can be reduced since less instructional material is needed. Too few phases leads to lower performance. Although the costs here too are reduced, the reduction in learning precludes this option. These results on learning are, in our opinion, not exclusive for Law but can be applied to all domains/situations where whole learning tasks are not too simple and where it is possible to identify functionally relevant phases in order to apply phase optimization. A consideration specific to our situation is that the instructional material presupposed that all students had roughly the same prior knowledge and skills. A current trend in education is to develop personalized, student-centered instruction that takes differences in prior knowledge and skills into account. This clearly results in more instructional material, as optimization on the number of phases needs to be done for student-groups differing in prior knowledge and skills. Data on mental effort and time on task seems a workable solution for tailoring the instructional material "on the fly".

Results here point out several directions for future research. First, since optimization of the number of phases leads to better performance, other instructional measures such as the availability of within-phase support could also be of importance. The next chapter reports on experiments on the relation between the number of phases and within-phase support to gain more insight into this. A second research thrust is needed with respect to the effects of the number of phases on transfer task performance and efficiency. In this study, the high-variability practice learning tasks were not compulsory and the participants could not be expected to invest more than 60 study hours for the Practical (the course containing the practical was a 100-hour module). Various studies have shown the benefits of a varied set of learning tasks on transfer (Cormier & Hagman, 1987; Paas & van Merriënboer, 1994; Quilicy & Mayer, 1996; Shapiro & Schmidt, 1982; Singly & Anderson, 1989). Whole learning tasks that take less time on task could bypass this practical obstacle, though there would be a risk of choosing whole tasks that are too simple. Finally, further research is needed to investigate whether the sort of learner control influences the support given.

Chapter 5 - Process support in learning tasks for complex cognitive skills*

Whole tasks for acquiring complex skills are often too difficult for novice learners. The common solution is to divide the problem solving process into phases, provide driving questions to help to carry out each phase, and give feedback at the end of each phase. Sophomore Law students ($N = 82$) participated in a multimedia training program designed to teach them to carry out a plea in court. In a 2 x 2 factorial design, in which the number of phases and the availability of driving questions were varied, the results indicate that students solving a *whole task* with less phases outperformed students exposed to more phases and also did this more efficient. Furthermore, students receiving driving questions outperformed those not receiving them, although they did not do this in a more efficient way.

Introduction

In the previous chapter, the effects of segmenting a whole learning task into smaller task assignments thereby dividing the problem solving process into *phases*, was presented as a way of pedagogically modeling a model (Achtenhagen, 2001). In this chapter, a second mechanism for process support is added, namely providing *driving questions* to guide carrying out the phases (Land, 2000). Driving questions are given at the start of a phase and guide the learners in how to carry out the activities within a phase, for instance, by referring them to information resources in a correct and efficient way, by activating relevant prior knowledge, or by suggesting relevant procedures and principles. In this approach, the process worksheet focuses on the problem solving process of the whole learning task while driving questions focus on the problem solving process within the phases. Both kinds of *process support* are domain-specific (see, e.g., Chinnappan & Lawson, 1996), thus tuned to the task at hand as opposed to being content-free heuristics (i.e., general problem solving methods) or content-free questions (e.g., in the case of writing a report: asking if the person checked the spelling, or added an index). Each of the process support mechanisms, can either separately or in combination, affect the effectiveness of the competency-based multimedia practical as reflected in the quality and/or efficiency of the task performance. Task efficiency is defined as task performance in relation to a combination of the mental effort, time on task, and motivation necessary to reach this level of performance.

Through tailoring the number of phases, complex learning tasks come within reach of learners' capabilities. Whole tasks or tasks with too few phases are often too difficult and too mentally demanding for learners, preventing them from accurately processing the necessary information because they experience cognitive overload or revert to superficial, non-meaningful learning to keep their cognitive load within the threshold limit (Craik & Lockhart, 1972; Sternberg & Frensch, 1991). Tasks with too many phases may also hamper learning because of their non-coherency as a result of redundant information between phases and/or an excess of details which makes them too mentally demanding (Mayer & Moreno, 2002). In addition, learners may regard the many phases as much too specific for the learning task in question, preventing them from constructing the generalizations or more abstract

---

* Based on: Nadolski, R. J., Kirschner, P. A., & van Merriënboer, J. J. G.. Process support in learning tasks for complex cognitive skills. Manuscript submitted for publication.

cognitive schemas necessary for learning transfer. In the previous study, we found that the number of phases in a whole learning task affected task-performance such that students receiving an intermediate number of phases outperformed students receiving either a low or a high number of phases (see Chapter 4).

Driving questions scaffold the problem solving process within the phases of whole learning tasks. In this study, the driving questions primarily guided learners in how to use appropriate resources and how to select relevant elements from these resources. They were provided at the start of the phase and kept available during task execution. They are expected to result in higher learner performance because they focus on higher order skills essential in problem solving (see, e.g., Smith & Ragan, 1999). In many respects, such questions are similar to *adjunct* questions which are "questions added to instructional text to influence what is learned from the text" (Hamaker, 1986, p. 212) and, in particular, similar to those adjunct questions which focus on higher order skills such as problem solving. The beneficial effects of adjunct questions have been consistently demonstrated (Hamaker, 1986; Hamilton, 1989, 1992; Peverly & Wood, 2001). Because both driving questions and higher-order adjunct questions focus on essentials in given information and its application, their effects are expected to be very similar.

In whole-task performance, an interaction effect might occur between the number of phases and the availability of driving questions. For a low number of phases, the problem solving process associated with the large phases can be expected to be (too) difficult when driving questions are not provided. Driving questions may have a positive effect on completing the phases in this situation, and thus on the performance for the whole task. For a high number of phases, the problem solving process for each phase will be simpler, so that the completion of the phases can be expected to be manageable without driving questions. As a consequence, if the whole task is divided in a low number of phases, providing driving questions is expected to be beneficial to the problem solving process for each phase and to whole-task performance; if the whole task is divided in a high number of phases, giving driving questions might provide too much process support for each phase, to the eventual detriment of whole-task performance.

The two mechanisms of process support can also influence task efficiency, that is, task performance in relation to the investments made to reaching this performance. Although efficiency can be operationalized in many different ways (see Admiraal, Wubbels, & Pilot, 1999; Kalyuga, Chandler, & Sweller, 1998; van Merriënboer, Schuurman, de Croock, & Paas, 2002), higher efficiency always indicates equivalent results with lower investments, higher results with the same investments, or, ideally, higher results with lower investments. In this study, investment is measured in terms of task motivation (Bonner, 1994; Maynard & Hakel, 1997), mental effort (Paas & van Merriënboer, 1994), and time on task (Karweit, 1984). In the previous study (see Chapter 4), the number of phases influenced task efficiency; with a high number of phases being least efficient. Participants in the condition with a high number of phases needed significantly more time on task than participants in both other conditions (low number of phases, intermediate number of phases). No differences between the conditions with respect to mental effort and motivation were found. It was expected driving questions to have either positive or neutral effects on motivation and mental effort because they act to demonstrate the relevance of the expected outcome to the learners. This perceived relevance may increase motivation and facilitate the process for carrying out the task which, in turn, may decrease the investment of mental effort (i.e., have a positive effect on mental effort). Furthermore, it was expected that the necessary time for answering driving questions counterbalances the extra time needed for finding the information when no questions are given. In other words, driving questions should not result in longer time on task and thus be advantageous to efficiency.

The present study employs a 2 x 2 factorial design to examine the effects of the number of phases (low, high) and driving questions (questions, no questions) on the performing tasks and on the efficiency of that performance. The first hypothesis is that students solving a whole learning task with a low number of phases will show higher performance and be more efficient learners than students exposed to a high number of phases. The second hypothesis is that students receiving driving questions for each phase will show higher performance and be more efficient learners than students not receiving driving questions. A third, more explorative, hypothesis pertains to a possible interaction between the number of phases and the availability of driving questions namely that driving questions have little added value if the number of phases is high, but become more valuable if the number of phases is low.

Method

Participants

Eighty-two sophomore Law students from five Dutch universities (49 female, 33 male; mean age = 23.5 years, *SD* = 4.2) enrolled in the experiment and were randomly assigned to one of four conditions. The conditions were: low number of phases without driving questions (*n* = 21); low number of phases with driving questions (*n* = 18); high number of phases without driving questions (*n* = 22); high number of phases with driving questions (*n* = 21). None of the participants had prior plea experience.

Learning materials

The Multimedia Practical *Preparing a plea* (Wöretshofer et al., 2000) was adapted for this experiment. Four versions of the practical were produced for the different experimental conditions. For all versions, objective task complexity of the assignments (1 = very simple, 2 = simple, 3 = complex, 4 = very complex) was determined by 32 different participants using the task-complexity instrument developed by Nadolski, Kirschner, van Merriënboer, and Wöretshofer (in press). The mean objective complexity for the tasks were 2.7 (*SD* = .3) for the Low number of phases – no driving questions, 2.3 (*SD* = .3) for the Low number of phases – driving questions, 2.0 (SD = .4) for the High number of phases – no driving questions, and 1.9 (*SD* = .4) for the High number of phases – driving questions. All versions contained identical information and support tools. The goal of the practical is to prepare students to carry out a plea in court. Figure 5.1 provides an overview of the learning materials for the four conditions. Participants receive a whole-task training using three learning tasks, one compulsory and two additional, non-compulsory learning tasks, as well as one transfer task. The support tools included modeling examples of lawyers conducting a plea, a "plea checker" to help them analyze pleas, discussions of ethical issues in pleading a case, numerous tips for communicative aspects in pleading a case, and judicial-procedural aspects of plea preparation. All versions contained case files and legal documentation (i.e., sections of law codes, jurisprudence), which was available to participants while working on the learning tasks. The material for all learning tasks also contained process worksheets that differed for each condition. Final performance on the whole learning task (i.e., the plea) is considered proof of skill acquisition. For all learning tasks, all conditions included condition specific feedback provided at the end of each phase.

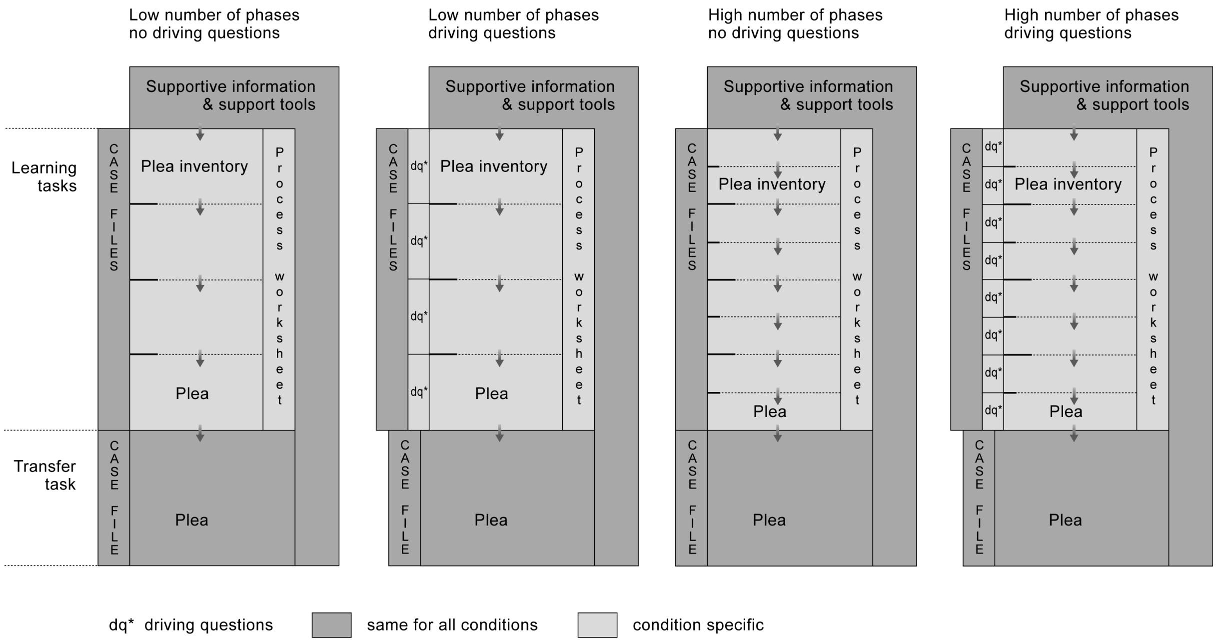


Figure 5.1
Overview of the learning materials in the four conditions.

Each phase included in the process worksheet was accompanied by a task assignment, with or without driving questions. One of the phases results in a *plea inventory*: these are the selected elements from the case file and legal documentation that might be usable in a plea. Table 5.1 presents the driving questions for the plea inventory assignment. The phase in which a plea inventory had to be constructed was included in all conditions. In the two conditions with a low number of phases it was the first phase; in the two conditions with a high number of phases it was the second phase. The material for the transfer task was the same in all conditions and included neither phases, driving questions, or feedback. Final performance on the transfer task (i.e., the plea) is considered proof of skill acquisition.

Table 5.1
Driving questions for 'Get acquainted with a law case': The answers result in a plea inventory

What field of law does this case belong to?
Who are the parties in this case?
Whose representative will you be?
Who is bringing the case before court?
What kind of procedure should be used?
What kind of judge will try this case?
What about the territorial jurisdiction of the judge?
What is this case roughly about?
Is there a previous judicial/legal history, and if so, is it relevant for this case?
What do you already know about the subject matter in this case?
Is the file complete?
What legal aspects could be of importance in this case?
What clues does the case offer at first sight?
What is the tentative goal of your plea?

Note that these questions are not specific to a certain case, but are specific to a certain domain

Measurement instruments

Backgrond questionnaire.

A background questionnaire gathered participants' data on prior plea experience (e.g., prior written and oral presentation skills, membership of debating club), computer literacy, attitude towards learning with computers, age, and gender.

Performance instrument for whole learning task.

An instrument (7-point-scale) was used to measure the performance results of participants' pleas on the compulsory learning task (i.e., learning plea). Inclusion of relevant content was measured by 39 items, while 7 items measured the coherence of the learning task plea (e.g., 'starts with a conclusion to be drawn from the plea', 'provides a content overview of the plea before going into details', 'indicates how successive arguments relate to relevant facts', 'various arguments are treated separately'). This instrument was the same as in the previous study, but dimensions relevant content and coherence were now collapsed into one performance dimension. The performance for the learning task plea was scored on those 46 items (Cohen's Kappa = .6).

Performance instrument whole transfer task.

An instrument (7-point scale) was used to measure the performance results of participants' transfer task pleas. The items were similar to those mentioned in the former instrument. Inclusion of relevant content was measured by 46 items, while 9 items measured the coherence of the transfer task plea. This instrument was the same as in the previous study, but dimensions relevant content and coherence were now collapsed into one performance dimension. The performance for transfer task was scored on those 55 items (Cohen's Kappa = .6).

Performance instrument for plea inventory.

An instrument (4-point scale) was developed to specifically measure the participants' performance on the plea inventory for the learning task (i.e., the inclusion of relevant content). The performance for the plea inventory was scored on 38 items (Cohen's Kappa = .7).

Instrument for time on task.

Participants reported time on task for each phase in multiples of five minutes on a pre-structured time sheet.

Mental effort rating scale.

Participants also indicated their mental effort for each phase on a 9-point rating scale (1 = very, very low mental effort, 9 = very, very high mental effort), developed by Paas, van Merriënboer and Adam (1994). This scale was used to measure the experienced cognitive load of each of the constituting phases of the learning task, and the cognitive load of the transfer task. Experienced mental effort was used as an indication for cognitive load; the lower the experienced cognitive load, the less mentally demanding the task.

Motivation rating scale.
In addition, participants indicated their motivation for each phase on a 3-item, 7-point rating scale (1 = very, very low, 7 = very, very high motivation), developed by Maynard and Hakel (1997). Items were: "I was motivated to perform well on this task assignment", "This task assignment was interesting to me", and "I put a lot of effort into coming up with the best possible solution".

Satisfaction/perceived efficacy rating scale.
Participants filled out rating scales to indicate their satisfaction with the quality of the instructional materials and the perceived efficacy of those materials. Satisfaction with feedback was indicated on an 8-item, 4-point scale (1 = very dissatisfied, 4 = very satisfied), the adequacy of the number of phases was indicated on a single 7-point scale (1 = far too few phases, 4 = perfectly all right, 7 = far too many phases), whether the goal of the practical ("I've learned to conduct a plea") was reached was scored on a single 7-point scale (1 = very strongly disagree, 7 = very strongly agree), and the relevance of the supportive information was indicated on a single 4-point scale (1 = strongly irrelevant, 4 = strongly relevant).

Computer logging.
Finally, since all conditions were computer-supported, study behavior was logged and analyzed. Participants' logging files were electronically collected for this purpose.

Procedure
Before taking part, participants were informed about the practical (60 study hours in one month), the required prior knowledge and skills, and the fact that data would be gathered for scientific research. Participants were randomly assigned to one of the four experimental conditions and were required to work individually. All materials were sent to the participants' home addresses or could be collected at their faculty addresses. All versions of the learning materials first presented identical information and support tools. After that, participants could work on the learning tasks. Participants were strongly advised to work phase-by-phase because the program offered the possibility of skipping consecutive phases. Logging results indicate that the participants did not skip any phases. Within a phase, there was maximum learner control so that participants were free to decide if and when to consult phase-specific information and how long to work on the task assignment in a phase.

After two weeks (approximately 30 study hours), participants were required to make their plea for the compulsory learning task. This plea was videotaped for later evaluation. After working on the learning task(s), participants had access to the transfer task, which did not contain any support. About two weeks later, they were required to hold their transfer task plea, which was also videotaped. There were strict time constraints for the pleas. Since it is legally required for lawyers in the Netherlands to submit a plea note when pleading a case in court, participants were required to include a plea note (i.e., a memorandum of the oral pleading) for both tasks. Participants were required to return the completed background questionnaire, time sheet, and rating scales in a stamped self-addressed envelope and to electronically send their logging results one week after completion of the course. Participants were informed whether they did or did not earn the study/course credits.

Data analysis and scoring
The experimenters extracted the plea inventory from the logging results and analyzed logged study behavior. Two judges blindly and independently scored all participants' word processed plea inventories and their videotaped pleas using the performance measurement instruments.

All efficiency measures were calculated using a procedure that closely resembles the procedure described by Tuovinen and Paas (in press) for determining instructional condition efficiency. Efficiency was calculated as (P-E-T-M) / SQRT (4), where P = performance, E = mental effort, T = time on task, and M = motivation. The P, E, T, and M scores on all variables are standardized (the total mean was subtracted from each score and the result was divided by the standard deviation), giving *z*-scores for each variable. This score can be negative, in the case that the sum of normalized scores for mental effort, time on task and motivation is greater than the normalized score for the performance.

Results

The collected data for determining computer literacy and attitude towards learning with computers showed no differences between conditions. None of the participants had prior plea experience, which was concluded from the collected background questionnaire data for prior plea experience. Participant comparability with respect to domain knowledge was assured since first year Law curricula of all Dutch universities are virtually identical with respect to both courses taken and textbooks used. Analysis of logging results indicated that participants did not spend any time on the non-compulsory learning tasks.

Performance

The mean performance results for the learning task, the transfer task, and the plea inventory are summarized in Table 5.2.

Table 5.2
Performance for the learning task, transfer task, and plea inventory

	Low number of phases				High number of phases			
	No questions (n = 21)		Questions[a] (n = 18)		No Questions (n = 22)		Questions (n = 21)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Learning plea*(1-7)	6.19	.87	5.94	.73	5.41	.96	5.33	1.24
Transfer plea (1-7)	3.29	.85	3.39	.61	3.19	.51	3.14	.73
Plea inventory**(1-4)	1.90	.62	2.38	.50	1.86	.56	2.57	.68

* $p < .01$ for the number of phases **$p < .001$ for availability of driving questions
[a] Only 16 of the 18 plea inventory results were collected for this condition.

With regard to the learning task, ANOVA revealed a main effect for the number of phases on the performance of the learning plea, $F(1, 78) = 10.39$, $MSE = .95$, $p < .01$, $\eta^2 = .12$. The conditions with less phases ($M = 6.08$, $SD = .16$) significantly outperformed the conditions with more phases ($M = 5.37$, $SD = .15$). There was no main effect for driving questions and no interaction.

With regard to the transfer task, results on the performance of the transfer plea revealed neither significant main effects nor interaction effects.

With regard to the quality of the plea inventory, ANOVA revealed a main effect for driving questions, $F(1, 76) = 19.02$, $MSE = .36$, $p < .001$, $\eta^2 = .20$. The conditions with driving questions ($M = 2.47$, $SD = .61$) significantly outperformed the conditions without driving questions ($M = 1.88$, $SD = .59$). There was no main effect for the number of phases and no interaction.

Time on task, mental effort, motivation

The mean results for time on task, mental effort, and motivation are summarized in Table 5.3.

Table 5.3
Time on task (in minutes), mental effort and motivation on learning task, transfer task, and plea inventory

	Low number of phases				High number of phases			
	No questions (n = 21)		Questions (n = 18)		No Questions (n = 22)		Questions (n = 21)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Learning task								
- time on task	933	274	947	348	1048	453	1170	446
- mental effort (1-9)	5.77	1.03	5.35	1.15	5.54	.94	5.45	.88
- motivation (1-7)	5.34	.83	5.29	.79	5.51	.70	5.64	1.00
Transfer task								
- time on task	351	177	360	174	425	309	411	213
- mental effort	5.70	1.74	5.42	1.27	6.20	1.45	5.48	1.41
- motivation	5.45	.78	5.57	.88	5.89	.69	5.64	.84
Plea inventory								
- time on task	210	84	208	115	175	103	172	77
- mental effort	5.19	1.57	5.00	1.88	5.14	1.32	5.10	1.22
- motivation	5.57	.82	5.46	.98	5.61	.86	5.41	.89

Time on task based upon self-report (reported in multiples of five minutes).
Mental effort is measured on a 9-point scale (Paas, van Merriënboer, & Adam, 1994).
Motivation is measured on a 3-item 7-point scale (Maynard & Hakel, 1997).

With regard to the learning task, there was a marginally significant effect for the number of phases on time on task, $F(1, 78) = 3.81$, $MSE = 152045$, $p < 0.1$, $\eta^2 = .05$. The conditions with less phases spent less time on task (M = 940 min, SD = 306) than conditions with more phases (M = 1109 min, SD = 449). There were no effects for the number of phases on mental effort and motivation. There were also no main effects for driving questions, and no interaction effects on time on task, mental effort and motivation.

With regard to the transfer task and the plea inventory, there were no significant main effects and no interaction effects on either of the dependent variables.

In sum, the analyses of variance revealed neither significant main effects nor interaction effects on mental effort or motivation. Participants reported an average mental effort, ranging between 5.00 and 6.20 (max = 9) for all three tasks in all conditions. Participants in all conditions were highly motivated when working on the learning task or the transfer task, and preparing their plea inventory, with motivation scores ranging between 5.29 and 5.89 (max = 7).

Task efficiency

There were no significant correlations between mental effort, motivation, and time on task. The mean efficiency results for learning task, learning plea, and plea inventory are summarized in Table 5.4.

Table 5.4
Efficiency for the learning task, transfer task, and plea inventory

	Low number of phases				High number of phases			
	No questions (n = 21)		Questions (n = 18)		No Questions (n = 22)		Questions (n = 21)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Learning task*	2.81	7.41	2.59	8.94	-0.43	11.64	-3.58	11.17
Transfer task	1.23	6.40	1.22	6.13	-1.63	10.95	-0.79	7.15
Plea inventory	-1.24	4.57	-0.72	6.43	.58	5.27	1.28	4.14

* p < .05 for the number of phases

With regard to learning task efficiency, ANOVA revealed a main effect for the number of phases, $F(1, 78) = 4.51$, $MSE = 99.92$, $p < .05$, $\eta^2 = .06$. The conditions with a lower number of phases ($M = 2.71$, $SD = 8.04$) were more efficient than the conditions with a higher number of phases ($M = -1.97$, $SD = 11.38$). There was no main effect for driving questions and no interaction effect.

With regard to the transfer task efficiency, there were neither significant main effects nor interaction effects.

With regard to plea inventory efficiency, ANOVA did not, contrary to the expectations, reveal a significant main effect for driving questions. There was no main effect for the number of phases and no interaction.

Satisfaction/ perceived efficacy

With regard to user satisfaction with feedback, there was no main effect for number of phases and no interaction (see Table 5.5).

Table 5.5
Satisfaction and perceived efficacy of the training

	Low number of phases				High number of phases			
	No questions (n = 21)		Questions (n = 18)		No Questions (n = 22)		Questions (n = 21)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
feedback[a]*	3.05	.44	3.24	.41	3.18	.42	3.38	.40
phases adequacy[b]	5.20	1.32	4.78	.94	5.18	.73	5.33	.97
goal of the practical[c]	5.52	.93	5.39	1.14	5.72	.94	5.57	.87
supportive info[d]	2.75	.55	2.61	.70	2.95	.49	2.76	.70

* p < .05 for the availability of driving questions
[a]Satisfaction with feedback (1 = very dissatisfied, 4 = very satisfied).
[b]The adequacy of the number of phases (1 = far too few, 4 = perfectly all right, 7 = far too many).
[c]Whether the goal of the Practical was reached; the assertion "I've learned to conduct a plea"(1 = very strong disagree, 7 = very strongly agree).
[d]The relevance of the supportive information (1 = strongly irrelevant, 4 = strongly relevant).

ANOVA indicated a main effect for driving questions, $F(1, 78) = 4.24$, $MSE = .18$, $p < .05$, $\eta^2 = .05$. Participants receiving driving questions ($M = 3.31$, $SD = .41$) reported that they were significantly more satisfied with feedback than participants receiving no driving questions ($M = 3.12$, $SD = .43$).

With regard to participants' opinion on the adequacy of the number of phases, their opinions on efficacy ("I've learned to conduct a plea"), and their opinions on the relevance of the supportive information, ANOVAs indicated neither significant main effects nor interaction effects.

Discussion

This study examined the effect of the number of phases and the availability of driving questions on both task performance and task efficiency. With regard to the effect of the number of phases, the results show a main effect for the number of phases on learning task performance as well as efficiency. A lower number of phases in learning to solve complex whole tasks led to both higher performance and greater efficiency. There were no differences between the conditions for transfer task performance and efficiency. These findings are in line with the earlier study in which only the number of phases was varied (see Chapter 4); where the low number of phases in this study is comparable to the intermediate number of phases in the previous one and the high number of phases in this study is the same as in the previous one. The finding that transfer task performance and efficiency were not influenced by the number of phases is most probably due to the fact that participants did not work on the non-compulsory learning tasks, which were meant to offer variability of practice and so induce transfer of learning. Thus, the learning situation was limited to the compulsory learning task (i.e., one single, whole learning task). Researchers agree that transfer cannot be expected under such circumstances (Gagné, Yekovich, & Yekovich, 1993; van Merriënboer, 1997).

With regard to the effect of driving questions on task performance and efficiency, the results show a positive main effect for driving questions on the plea inventory performance, which is one of the phases in the whole learning task. The availability of driving questions did not - contrary to the expectations - beneficially affect the efficiency of the plea inventory due to an unexpected higher time on task. Although participants exposed to more phases showed - as expected - a trend towards spending more time on the whole learning task as opposed to those exposed to less phases, an opposite trend was found for time on task on the plea inventory. This is possibly due to the position of the plea inventory assignment in the whole set of instructional materials. In conditions with fewer phases the plea inventory is the first assignment, but in conditions with more phases it is the second assignment. Participants receiving the plea inventory as the first assignment might be inclined to invest more time than those receiving it as the second assignment to get acquainted with the whole task. The absence of an effect of driving questions on both learning task performance and efficiency is possibly due to task characteristics and feedback given at the completion of phases. In this study, the problem solving process can be seen as a sequence of interdependent phases converging towards a solution for the whole learning task. In fact, each subsequent phase further decreases the problem space for the whole task, and feedback is provided after each phase. Therefore, positive effects of driving questions are expected to extinguish in subsequent phases. This explanation, in addition to the observation that participants did not use the non-compulsory learning tasks, probably also accounts for driving questions not positively affecting transfer task performance and efficiency.

Finally, no interaction effects were observed between the number of phases and the provision of driving questions. There are two possible explanations for this. First, the difference between the low and the high number of phases might have been too small, which prevented the expected added value of driving questions to occur. Second, the decreasing problem space and the provision of feedback after completion of the phases might also account for the absence of interaction effects.

The results imply several directions for future research. First, further experiments should offer a compulsory set of varied learning tasks for practice, making sure that participants work on a set of learning tasks with a high variability. Various studies have shown the benefits of high-variability practice on transfer (Cormier & Hagman, 1987; Paas & van Merriënboer, 1994; Quilicy & Mayer, 1996; Shapiro & Schmidt, 1982; Singly & Anderson, 1989). Second, future experiments should further explore the conditions under which driving questions lead to better performance and efficiency. These conditions appear to be related to whole-tasks characteristics such as the size of the problem space (Newell & Simon, 1972), the kind of solution (convergent vs. divergent), the relationship between phases (interdependent vs. independent), and the provision of feedback after completion of the phases. For example, a study examining the effects of driving questions on the learning of generative and creative brainstorming tasks could provide more insight in the value of driving questions. Third, future studies must take into account that process support may be domain-related. It is expected that the reported findings can be extended to a domain with an ontology that is similar to that of Law. However, if the ontology of a domain is different from Law, it is unclear if the findings of the current study could be replicated. Finally, future research should maximize the difference between a low and high number of phases, which may make it possible to find an interaction effect between the number of phases and the availability of driving questions on task performance and task efficiency.

A straightforward practical implication of this study is that process support should be provided for whole learning tasks. One should split the whole task in a limited number of phases but not too many phases. Practical implications for the use of driving questions are less clear. In this study it was found that driving questions were probably only beneficial for early phases in problem solving, due to task characteristics and feedback given at the completion of each phase. Therefore, more articulated research with respect to driving questions is needed. A final consideration is that the instructional material used presupposes that all students have roughly the same prior knowledge and skills. However, mental effort and time on task may provide good input for tailoring the instructional materials to individual students while they are working with it. Such approaches to personalized, student-centered instruction not only provide whole tasks to students, but also offer the opportunity to build Multimedia Practicals that adapt their level of process support to individual learners.

Chapter 6 - General discussion

The main aim of this thesis was to determine the effects of process support on the acquisition of complex cognitive skills within Multimedia Practicals and so provide guidelines for designing such learning environments. The premise was that a whole-task approach should be used for the acquisition of complex cognitive skills, but since whole tasks are often too difficult for novice learners, the process of executing the task should be facilitated through process support. Process support was provided by (a) dividing the whole learning task into a number of smaller phases, which were specified in a process worksheet reflecting a systematic approach to problem solving, and (b) offering driving questions to help learners within the phases.

The central research questions addressed are:

(1) does the number of phases influence performance on the task and the efficiency of carrying out the task, and if so, in what way?

(2) do driving questions positively influence performance on the task and the efficiency of carrying out the task?

In this thesis a design approach has been developed and tested, a measurement instrument needed for that approach has been developed and validated, and the central research questions have been studied. This final chapter provides a review of the design approach and discusses the results of the empirical studies, presents practical implications and guidelines of the research, and gives suggestions for future research. It concludes with a final remark on the ecological validity of the presented studies.

Design approach and results

Design approach

In Chapter 2 the application of a two-phase six-step Instructional Design model for Multimedia Practicals resulted in a detailed blueprint for their effective development. Carrying out the design phase is largely based on the 4C/ID-model (van Merriënboer, 1997) and insights from cognitive load theory (Sweller, 1988). It results in an instructional sequence of whole tasks, with process support for these tasks in order to facilitate the acquisition of the complex cognitive skill that is taught. The Instructional Design model requires the measurement of the objective complexity of learning tasks. The measurements of task complexity are needed to determine the number of phases in a systematic approach to problem solving, which is subsequently presented to the learners as a process worksheet.

Task complexity instrument

A four-category scale instrument for objectively determining learning task complexity was developed as a component of the design approach (Chapter 3). This benchmark instrument, with anchor tasks representing the four categories of task complexity, proved to be reliable, easy to use, and learnable without specific training. Studies on the use of this instrument showed that advanced level Law graduate students were able to reliably determine task complexity at a level comparable to that of experienced Law teachers. This simplifies and reduces the cost of designing and developing Multimedia Practicals. The instrument is powerful for determining the complexity of learning tasks at the extremes of the scale, but is somewhat less powerful for determining intermediate levels of complexity.

Answering the research questions

The first study (Chapter 4) examined the effects of the number of phases. It showed that limiting the number of phases in learning to solve complex tasks leads to optimal task performance. Too many phases appeared to make the learning task less coherent. Though time on task was increased for the learner, no concomitant increase in performance was observed. Too few phases, while as efficient as the optimal condition with respect to amount learnt per unit of time, resulted in a lower performance on the learning task. The second study (Chapter 5) examined the effects of the number of phases and the use of driving questions on both task performance and task efficiency, and replicated these results. Again, a main effect was found for the number of phases, showing that limiting the number of phases in learning to solve complex tasks leads to higher task performance. The low number of phases in the second study was comparable to the intermediate number of phases in the first study, and the high number of phases in the second study was the same as in the first study. In both studies, as predicted, *learning efficiency* was lowest for learners confronted with a high number of phases. Learners in the conditions with many phases showed a significantly higher time on task, without showing improved learning. Contrary to the expectations, neither differences on time on task or on task efficiency were found between learners in the Single phase and Intermediate number of phases conditions in the first study. In the second study, however, a low number of phases led to greater task efficiency, showing a main effect for the number of phases.

Both studies showed, contrary to the expectations, that transfer task performance and transfer task efficiency were not influenced by the number of phases. This is most probably due to the fact that participants did not work on the non-compulsory extra learning tasks, which were meant to offer variability of practice and so induce transfer of learning. The learning situation was thus limited to the compulsory learning task (i.e., one single, whole learning task). Researchers agree that transfer cannot be expected under such circumstances (Gagné, Yekovich, & Yekovich, 1993; van Merriënboer, 1997).

The second study also shed light on the second research question, namely whether the availability of driving questions positively influences task performance and task efficiency. Indeed, a positive main effect was found for driving questions on plea inventory performance, one of the phases in the whole learning task. However, the availability of driving questions did not beneficially affect the efficiency of the plea inventory. This is possibly due to an unexpectedly higher time on task resulting from the position of the plea inventory in the whole set of instructional materials. The absence of an effect of driving questions on both learning task performance and learning task efficiency is probably due to task characteristics and feedback given at the completion of phases.

Finally the second study explored whether an interaction effect between the number of phases and the provision of driving questions occurred on task performance and efficiency. It was expected that driving questions would have little added value if the number of phases were high (i.e., each phase is so simple that it can be easily completed without help from driving questions), but would become more valuable if the number of phases is low. No interaction effects were observed. Two possible explanations are that (1) the difference between the low and the high number of phases was too small, which prevented the expected added value of driving questions to occur and (2) the process support was set up in such a way that driving questions quickly lost their added value, because the problem space became smaller and smaller for later phases and feedback was provided after completion of each successive phase.

Practical implications and guidelines

The design of Multimedia Practicals is a costly and time-consuming process. Development and eventual maintenance is even more costly when the design is less detailed and consistent (see, e.g., Sherlund, Wade, Emery, & Hilliard, 2000). Although the application of the two-phase six-step Instructional Design model for Multimedia Practicals is resource-intensive in the cognitive task analysis phase, it should ultimately save costs during development and result in better maintainable, higher-quality materials. In addition, the study on the task complexity instrument shows that determining objective task complexity prevents serious mistakes in the analysis and design phases, and also lowers the costs as the pool of consultants for using this instrument increases, that is, advanced level graduate students can be used as raters because they rate similarly to experienced teachers. The Instructional Design model used for this research thus clearly differs from other task-analytical methods, which do not consider a measurement of objective task complexity.

A major implication of the work presented is that instruments are needed for the measurement of objective task complexity, so that an appropriate number of phases can be determined to optimize process support. This research shows that it is possible to develop an instrument that allows developers to measure task complexity at reasonable costs prior to confronting learners with the learning environment and thus to design the learning environment in accordance with the complexity of the tasks. Although the instrument described in this thesis is dedicated to measuring the complexity of Law learning tasks for sophomore Law students, its conceptual frame of reference offers a firm starting point for developing analogous instruments for other learning domains.

The practical implications of the two empirical studies on process support are quite straightforward. There is clear empirical evidence for the value of offering process support through the distinction of a number of phases. Both studies showed that it is recommendable not to use too many phases. Too many phases leads to lower performance and does not justify the extra costs for developing such detailed instructional materials. Put bluntly, development costs can often be reduced since less instructional material is needed. But, the ultimate reduction of the number of phases to a Single-phase whole task, as seen in the first study, also leads to lower performance. Thus, a further reduction in costs is precluded because a moderate number of phases should be distinguished to reach acceptable learning performance.

The finding that driving questions were probably only beneficial for early phases in problem solving, due to task characteristics and feedback given at the completion of each phase, implies that such driving questions might be useful, but that they can and should be faded during the learning process. Finally, the findings can possibly be extended to domains with ontologies similar to that of Law (i.e., argumentative and dialogical as in philosophy and mathematics). However, if the ontology of a domain is different from Law (i.e., inquiry-based as in science, or design-based as in engineering), it is unclear if the findings of this research project could be replicated. Task characteristics such as the size of the problem space, the kind of solution (convergent vs. divergent), the relationship between phases (interdependent vs. independent), and the provision of feedback after completion of the phases all influence the effects of process support.

In sum, these practical implications lead to the following design guidelines for process support in learning environments for the acquisition of complex cognitive skills:

a. determine the objective complexity of learning tasks
b. provide process support that takes task complexity into account
c. split the problem solving process of the whole task into phases, but not too many
d. consider providing driving questions in the early phases of problem solving
e. do all of this on the basis of a systematic design methodology, such as the two-phase six-step Instructional Design model for Multimedia Practicals.

Further research

The empirical studies on process support provide a start for justifying the Instructional Design model for Multimedia Practicals described in Chapter 2. More research and development should be directed towards a further specification and articulation of the model, necessary for making it directly useful for less experienced designers or teachers. This is in line with formative research on the simplifying conditions method, which also aims at increasing the usability of design methodologies (Reigeluth, Lee, Peterson, & Chavez, 1999).

The promising results on the task complexity measurement instrument, however, need to be qualified. First, extremes proved to be easier to define than centralities. Second, further refinement is possible. Because complexity is a multidimensional concept, one may question whether the developers' one-dimensional approach of complexity concentrating only on intellectual operations is sufficient. For objectively measuring the complexity of learning tasks, two promising lines for further research can be distinguished. First, new studies must make the conceptual frame of reference as well as the way that anchor tasks fit within this frame explicit, to investigate if this results in more confident ratings for – especially - the intermediate complexity tasks. Second, new studies must determine the added value of a multidimensional construct of learning task complexity above the one-dimensional construct used in this thesis (i.e., intellectual operations), using dimensions such as quantity of information searched for and combined, field or discipline of learning, symbolic system of task-formulation (text and/or graphics, animations), and preferences and learning styles of the receiver.

The results of the two studies on process support also point out several directions for future research. First, a follow-up experiment should offer a *compulsory* set of high-variability learning tasks for practice. Various studies have shown the benefits of a varied set of learning tasks on transfer (Cormier & Hagman, 1987; Paas & van Merriënboer, 1994; Quilicy & Mayer, 1996; Shapiro & Schmidt, 1982; Singly & Anderson, 1989). In our studies, the participants could not be expected to invest more than 60 study hours for the practical, which was part of a regular 100-hour module. Whole learning tasks that take less time on task could bypass this practical obstacle, though there would be a risk of choosing whole tasks that are too simple. Second, future experiments should further explore the conditions under which driving questions lead to higher performance and efficiency. These conditions appear to be related to whole-task characteristics such as the size of the problem space (Newell & Simon, 1972), the kind of solution (convergent vs. divergent), the relationship between phases (interdependent vs. independent), and the provision of feedback after completion of each phase. For example, a study examining the effects of driving questions on the learning of generative and creative brainstorming tasks (large problem space, divergent solutions) could provide more insight in the value of driving questions. A third line of research should try to replicate the reported results in other domains with both similar and different ontologies. Fourth, research is needed to investigate whether learner control influences the process support given. In our experiments, process support was provided at the start of problem solving, remained available during problem solving (students could have a look at it at any time), and did not change over time. Providing process support on students' request could be a better way of supporting learning than providing it at predetermined points in time (i.e., timing of process support, see Hummel & Nadolski, 2002; van Merriënboer, Kirschner, & Kester, 2003). Another consideration on learner control is that the instructional materials used in the studies presupposes that all students have roughly the same prior knowledge and skills. However, data on mental effort and time on task may provide good input for tailoring the instructional materials to individual students in a dynamic fashion, that is, while they are working with the materials. Such approaches to personalized, student-centered instruction not only provide whole tasks to students, but also offer the opportunity to build Multimedia

Practicals that adapt their level of process support to individual learners. Finally, fine-grained research is needed in less ecologically valid settings so as to maximize the difference between a low and high number of phases, which may make it possible to find an interaction effect between the number of phases and the availability of driving questions on task performance and task efficiency.

In this way, acquiring complex cognitive skills within Multimedia Practicals can become more effective and efficient, and address our society's need for more employees having flexible problem solving behavior at their disposal.

Final remark

Both experiments on process support were carried out in an ecologically valid context. Optimization of results by inclusion of a 'poor' learning condition with no learner support at all (i.e., where one can be fairly sure that learning is suppressed) was not an ethical option. The learning materials that formed the starting point of this research could be considered to be of high quality, and all the experimental conditions had the aim to 'make good materials even better'. This probably reduced the experimental effects because it was found that participants in conditions without process support (i.e., the Single-phase condition in the first study on process support) also performed reasonably well on the learning tasks. It is defensible to expect that leaving out the basic support mechanisms would have induced stronger effects for process support. However, since the participants were regular students, working for study credits, this was not a real option.

References

Achtenhagen, F. (2001). Criteria for the development of complex teaching-learning environments. *Instructional Science, 29,* 361-380.

Admiraal, W. , Wubbels, T. , & Pilot, A. (1999). College teaching in legal education: Teaching method, students' time-on-task, and achievement. *Research in Higher Education, 40,* 687-704.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369-406.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of eductional objectives*. New York: Addison Wesley Longman.

Bandura, A. (1982). Self-efficacy mechanism in human accuracy. *American Psychologist, 37,* 122-147.

Bedard, J., & Chi, M. T. H. (1992). Expertise. Current directions. *Psychological Science, 1,* 135-139.

Boekaerts, M., & Simons, P. R. J. (2003). *Leren en instructie* [Learning and instruction] (3rd ed.). Assen, The Netherlands: Koninklijke Van Gorcum.

Boggs, D. H., & Simon, R. J. (1968). Differential effect of noise on tasks of varying complexity. *Journal of Applied Psychology, 52,* 148-153.

Bonner, S. (1994). A model of the effects of audit task complexity. *Accounting, Organizations and Society, 19*(3), 213-234.

Brandle (2002). *Sentence complexity*. Retrieved February 17, 2003, from http://www.brandle.com.au/sentence.htm

Brown, J. S., Collins, A., & Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18,* 32-42.

Burtch, L. D., Lipscomb, M. S., & Wissman, D. J. (1982). *Aptitude requirements based on task difficulty* (Report No. AFHRL-TR-81-34). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management, 31,* 191-213.

Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review, 13,* 40-52.

Campbell, D. J., & Gingrich, K. (1986). The interactive effects of task complexity and participation on task performance: A field experiment. *Organizational Behaviour and Human Decision Processes, 38,* 162-180.

Chang, C. J., Ho, J. L.Y., & Liao, W. M. (1997). The effects of justification, task complexity and experience/training on problem-solving performance. *Behavioral Research in Accounting, 9,* 98-116.

Chase, W. G., & Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics knowledge by experts and novices. *Cognitive Science, 5,* 121-152.

Chi, M. T .H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-75). Hillsdale, NJ: Erlbaum.

Chinnappan, M., & Lawson, M. J. (1996). The effects of training in the use of executive strategies in geometry problem solving. *Learning and Instruction, 6,* 1-17.

Clark, R. (1998). *Building expertise: Cognitive methods for training and performance*. Washington, DC: International Society for Performance Improvement.

Clark, R. E. (1999). Yin and Yang cognitive motivational processes in operating multimedia learning environments. In J. J. G. van Merriënboer (Ed.), *Cognition and multimedia design* (pp. 73-107). Heerlen, The Netherlands: Open University of the Netherlands.

Clough, P. (2000). *Analysing style-readability*. Retrieved February 17, 2003, from http://www.dcs.sheef.ac.uk/~cloughie/papers/readability.pdf

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-493). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooke, N. J. (1994). Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies, 41,* 801-849.

Cormier, S. M., & Hagman, J. D. (Eds.). (1987). *Transfer of learning: Contemporary research and applications.* San Diego, CA: Academic Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Crombag, H. F., Chang, T. M., Drift, K. D. J. M. van der, & Moonen, J. M. (1979). *Onderwijsmiddelen voor de Open Universiteit Nederland: Functies en kosten* [Educational materials for the Open University of the Netherlands: Functions and costs]. The Hague, the Netherlands: Ministry of Education and Sciences.

De Block, A. (1975). *Taxonomie van leerdoelen* [Taxonomy of learning objectives]. Antwerp-Amsterdam: Standard Academic Publishers.

De Croock, M. B. M. (1999). *The transfer paradox: Training design for troubleshooting skills.* Unpublished doctoral dissertation. Enschede, The Netherlands: University of Twente.

Dehoney, J. (1995). Cognitive task analysis: Implications for the theory and practice of instructional design. *Proceedings of the annual national convention of the Association for Educational Communications and Technology (AECT), Anaheim, CA,* 113-123. (ERIC Document Reproduction Service No. ED 383 294)

Dick, W., & Carey, L. (1979). *The Systematic Design of Instruction*. Glenview, IL: Scott Foresman.

DuBois, D. A., Shalin, V. L., Levi, K. R., & Borman, W. C. (1995). *A cognitively-oriented approach to task analysis and test development.* Minneapolis, MN: Personal Decisions Research Institute.

Early, P. C. (1985). Influence of information, choice and task complexity upon goal acceptance, performance, and personal goals. *Journal of Applied Psychology, 70,* 481-491.

Edens, F. M., Rink, F., & Smilde, M. J. (2000). De studentenrechtbank: Een evaluatieonderzoek naar beoordelingslijsten voor presentatievaardigheden [The student law court: An evaluation study on assessmentlists for presentation skills]. *Tijdschrift voor Onderwijsresearch, 24,* 265-274.

Eraut, M. (1997). *Developing professional knowledge and competence.* London: The Falmer Press.

Flesch, R. (2003). Chapter 2: Let's start with the formula. In: *How to write plain English*. Retrieved February 17, 2003, from http://www.mang.canterbury.ac.nz/courseinfo/AcademicWriting/Flesch.htm

Gagné, E., Yekovich, C., & Yekovich, F. (1993). *The cognitive psychology of school learning.* New York: Harper Collins.

Gardner, M. K., (1985). Cognitive psychological approaches to instructional task analysis. In E. W. Gordon (Ed.), *Review of Research in Education* (Vol. 12, pp. 157-195). Washington, D.C.: American Educational Research Association.

Gerrichhauzen, J. T. G., Hoefakker, R. E., Perreijn, A. C., van den Brink, H. J., Slootmaker, A. , & Berkhout, J. (1998). *Buiten dienst* [Out of order] (version 1.0) [multimedia CD-ROM]. Heerlen, The Netherlands: Open University of the Netherlands.

Guilford, J. P. (1982). Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review, 89,* 48-59.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56,* 212-242.

Hamilton, R. (1989). The role of concept definition, teaching examples and practice on concept learning from prose. *Contemporary Educational Psychology, 14,* 357-365.

Hamilton, R. (1992). Application adjunct post-questions and conceptual problem solving. *Contemporary Educational Psychology, 17,* 89-97.

Hannafin, M., Land, S., & Oliver, K. (1999). Open learning environments: Foundations, methods, and models. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 115-140). Mahwah, NJ: Lawrence Erlbaum.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: CBS College Publishing.

Hoogveld, A. W. M., Vander Meeren, W. M. F., Hummel, H. G. K., Vos, M. M. H. L. S., Rosendaal, A., van der Vegt, G. W., & Berkhout, J. (1997). *Practical assessment center* [Practicum assessment center] (version 2.0) [multimedia CD-ROM]. Heerlen, The Netherlands: Open University of the Netherlands.

Hommes, M. A., Houtmans, M. A., Hummel, H. G. K., Kuntze, A. J. , Tiesnitsch, D., Rickhoff, M. L., Westera, W., Kerstjens, W. M. J., Kurvers, H. J., Slot, W. J. J., Vander Meeren, W. M. F., & Berkhout, J. (2000). *Diagnosticus* [Diagnostician] (version 1.0) [multimedia CD-ROM]. Heerlen, The Netherlands: Open University of the Netherlands.

Huber, V. L. (1985). Effects of task difficulty, goal setting, and strategy on performance of a heuristic task. *Journal of Applied Psychology, 70,* 492-504.

Hummel, H. G. K., & Nadolski, R. J. (2002). Cueing for schema construction: Designing problem-solving multimedia practicals. *Contemporary Educational Psychology, 27,* 229-249.

Huysse, P., Nadolski, R. J., Oldenboom, E., Kerstjens, W. M. J., de Vries, F. J. J, Slootmaker, A, Jordense, M., Jambos, M., & Berkhout, J. (1998). *Paradise Parks* (version 1.0) [multimedia CD-ROM]. Heerlen, The Netherlands: Open University of the Netherlands.

Ivens, W. P. M. F., Lansu, A. L. E., Hummel, H. G. K., Huisman, W. H. T., Westera, W. , Wagemans, L. J. J. M., Slootmaker, A., & Berkhout, J. (1998). *Bodem en Milieu* [Soil and environment] (version 1.95) [multimedia CD-ROM]. Heerlen, The Netherlands: Open University of the Netherlands.

Jelsma, O. (1989). *Instructional control of tranfer.* Enschede, The Netherlands: Bijlstra & van Merriënboer.

Jonassen, D. H. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 371-396). Mahwah, NJ: Lawrence Erlbaum Associates.

Jonassen, D. H., & Hannum, W. H. (1995). Analysis of task analysis procedures. In G. J. Anglin (Ed.), *Instructional technology, past, present, and future* (pp.197-214). Englewood, CO: Libraries Unlimited.

Jonassen, D. H., Hannum, W. H., & Tessmer, M. (1989). *Handbook of task analysis procedures.* New York: Preager.

Jonassen, D. H., Tessmer, M., & Hannum, W. H. (1999). *Task analysis methods for instructional design.* Mahwah, NJ: Lawrence Erlbaum Associates.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40,* 1-17.

Karweit, N. (1984). Time-on-task reconsidered: Synthesis of research on time and learning. *Educational leadership: Journal of the Association for Supervision and Curriculum Development, 41,* 32-35.

Kernan, M. C., Bruning, M. S., & Miller-Guhde, L. (1994). Individual and group performance: Effects of task complexity and information. *Human Performance, 7,* 273-289.

Kirschner, P. A. (1991). *Practicals in higher science education.* Utrecht, The Netherlands: Lemma.

Kirschner, P. A., van Vilsteren, P. P. M., Hummel, H. G. K., & Wigman, M. C. S. (1997). The design of a study environment for acquiring academic and professional competence. *Studies in Higher Education, 22*(2), 151-171.

Kolb, D. A. (1984). *Experiential learning experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice Hall.

Land, S. M. (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development, 48,* 61-78.

Larken, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science, 4,* 317-345.

Leinders, J. J. M., Drury, S. A., Rothery, D. A., Nadolski, R. J., van der Heijden, M. P., de Vries, F. J. J., Slot, W. J. J., Roosendaal, A., Kurvers, H. J., & Berkhout, J. (1993). *Reseat* (version 1.0) [multimedia CD-ROM]. Heerlen, The Netherlands: Open University of the Netherlands.

Lewy, A., & Báthory, Z. (1994). The taxonomy of educational objectives in Continental Europe, the Mediterranean and the Middle East. In L.W. Anderson & L.A. Sosniak (Eds.), *Bloom's taxonomy: A fourty-year retrospective* (pp. 146-163). Chicago, IL: The University of Chicago Press.

Mayer, R. E. (1999). Designing instruction for constructivist learning. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol 2., pp. 141-159). Mahwah, NJ: Lawrence Erlbaum.

Mayer R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12,* 107-119.

Maynard, D. C., & Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance, 10*(4), 303-330.

Merrill, M. D. (1983). Component display theory. In C. M. Reigeluth (Ed.), *Instructional design theories and models: An overview of their current status* (pp. 278-333). Hillsdale, NJ: Lawrence Erlbaum.

Merrill, M. D. (1987). The New Component Design Theory: Instructional Design for Courseware Authoring. *Instructional Science, 16,* 19-34.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology, Research and Development, 50,* 43-59.

Nadolski, R. J., Kirschner, P. A., van Merriënboer, J. J. G., & Hummel, H. G. K. (2001). A model for optimizing step size of learning tasks in competency-based multimedia practicals. *Educational Technology Research and Development, 49,* 87-103.

Nadolski, R. J., Kirschner, P. A., van Merriënboer, J. J. G., & Wöretshofer, J. (in press). Development of an instrument for measuring the complexity of learning tasks. *Educational Research and Evaluation.*

Nelson, L. M. (1999). Collaborative problem solving. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 241-267). Mahwah, NJ: Lawrence Erlbaum Associates.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach. *Journal of Educational Psychology, 86,* 122-133.

Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79,* 419-430.

Peverly, S. T., & Wood, R. (2001). The effects of adjunct questions and feedback on improving the reading comprehension skills of learning-disabled adolescents. *Contemporary Educational Psychology, 26,* 25-43.

Pikulski, J. J. (2002). *Readability*. Retrieved February 17, 2003, from htttp://www.eduplace.com/state/author/pikulski.pdf

Quilicy, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 28,* 175-224.

Reigeluth, C. M. (Ed.). (1983). *Instructional theories and models. An overview of their current status.* Hillsdale, NJ: Lawrence Erlbaum.

Reigeluth, C. M. (Ed.). (1987*). Instructional theories in action. Lessons illustrating selected theories and models.* Hillsdale, NJ: Lawrence Erlbaum.

Reigeluth, C. M. (Ed.). (1999). *Instructional design theories and models: A new paradigma of instructional theory* (Vol. 2). Mahwah, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M., Lee, J. Y., Peterson, B., & Chavez (1999). *Formative Research on the Heuristic Task Analysis Process*. Retrieved January 19, 2001, from http://www.indiana.edu/~idtheory/hta.html

Reigeluth, C. M., & Merrill, M. D. (1984). *Extended task analysis procedures (ETAP): User's manual.* Lanham, MD: University Press of America.

Resnick, L. B. (1976). Task analysis in instructional design: Some cases from mathematics. In D. Klahr (Ed.), *Cognition and Instruction* (pp. 51-80). Hillsdale, NJ: Lawrence Erlbaum.

Romiszowski, A. J. (1981). *Designing instructional systems: Decision making in course planning and curriculum design*. London: Kogan Page.

Roth, E. M., & Woods, D. D. (1989). Cognitive task analysis: An approach to knowledge acquisition for intelligent system design. In G. Guida & C. Tasso, (Eds.), *Topics in expert system design* (pp. 233-264). New York: Elsevier.

Scott, W. E., Fahr, J., & Podsakoff, P. M. (1988). The effects of "intrinsic" and "extrinsic" reinforcement contingencies on task behavior. *Organizational Behavior and Human Decision Processes, 41,* 405-425.

Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Kelso & J. E. Clark, (Eds.), *The development of movement control and coordination* (pp. 113-150). New York: John Wiley & Sons.

Sherland, B., Wade, R. L., Emery, D., & Hilliard, R. (2000). *IEEE recommended practice for architectural description for software-intensive systems*. New York: Institute of Electrical and Electronics Engineers.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Singley, M. K., & Anderson, J. R. (Eds.). (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Smith, P. L., & Ragan, T. J. (1999). *Instructional Design* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1988). *Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains* (Tech. Rep. No. 441). Champaign, IL: University of Illinois, Center for study of reading.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (1998). Instructional effects in complex learning: Do objective and subjective learning outcomes converge? *Learning and Instruction, 8,* 117-129.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J., & Frensch, P. A. (Eds.). (1991). *Complex problem solving: Principles and mechanisms.* Hillsdale, NJ: Lawrence Erlbaum.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12,* 185-233.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

Taylor, M. S. (1981). The motivational effects of task challenge: A laboratory investigation. *Organizational Behavior and Human Performance, 27,* 255-278.

Tennyson, R. D., Elmore, R. L., & Snyder, C. (1991, april). Advancements in Instructional Design theory: Contextual module analysis and integrated instructional strategies. In M. Rausch (Chair), *Instructional design theory contributions to the cognitive science.* Symposium conducted at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 336 080)

Tuovinen, J. E., & Paas, F. (in press). Exploring multidimensional approaches to the efficiency of instructional conditions. *Instructional Science.*

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills.* Englewood Cliffs, NJ: Educational Technology Publications.

Van Merriënboer, J. J. G., De Croock, M. B. M., & Jelsma, O. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a complex cognitive skill. *Perceptual and Motor Skills, 84,* 784-786.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2001). Three worlds of instructional design: State of the art and future directions. *Instructional Science, 29,* 429-441.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking of the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5-13.

Van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test and training efficiency. *Learning and Instruction, 12,* 11-37.

Van Parreren, C. F., (1987). *Ontwikkelend onderwijs* [Developing education]. Amersfoort, The Netherlands: ACCO.

Vaso, G. (2000). *Determining the readability of a book*. Eclectic Homeschool. Bellevue. Retrieved February 17, 2003, from http://www.eho.org/pdf/100ln.pdf

Westera, W., & Sloep, P. B. (1998). The Virtual Company: Toward a self-directed, competence-based learning environment in distance education. *Educational Technology, 38,* 32-37.

Wilson, B., & Cole, P. (1990). A review of cognitive teaching models. *Educational Technology Research and Development, 39,* 47-64.

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology, 89,* 397-410.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational behaviour and human decision processes, 37,* 60-82.

Wöretshofer, J., Nadolski, R. J., Starren-Weijenberg, A. M. A. G., Quanjel-Schreurs, R. A. M., Aretz, C. C. W. M., van der Meer, N. H. W., Martijn, G., van den Brink, H. J., Slootmaker, A., & Berkhout, J. (2000). *Preparing a plea* [Pleit voorbereid] (version 1.0) [multimedia CD-ROM]. Heerlen, The Netherlands: CIHO.

Appendixes

Appendix 1

Activities and outcomes when applying the two-phase six-step instructional design model

Step-description	*Activities (by instructional designer)*	*Outcomes*
Step 1—CTA phase *Skill decomposition*	Start an introductory reading or research in the task field. Determine expected task behavior and expected prior knowledge and skills for the inexperienced task-performer during (written & oral) communication with teacher experts.	Formulation of the global competency to be achieved after finishing instruction. Formulation of prior knowledge and skills for the task-doer.
Step 1a—CTA phase Skill decomposition: Segmentation analysis	Identify segments for the task at hand by: - observations of expected task behavior by juveniles and experts (reality, taped); if possible discuss those observations afterwards - (if available) further readings about task (to become familiar with the field) - structured interviews asking experts (with different backgrounds) how they (mentally) perform the task, how they did learn to perform the task (how long did it take them to learn it), what does their preparation consist of, to decribe some case studies they encountered, what in their opinion are easy/complex segments, if there are part-tasks that are not always encountered (and why so), does a simplest version of the task exist and how representative is it, et cetera. (in these interviews it is advised to start with a juvenile expert which does not have internalized too much so the instructional designer can become familiar with the task and can further develop his/her interview-technique) - structured interviews to (further) identify a possible order in which part-tasks are conducted and check this order with different experts - report on temporary identified segments to teacher-experts and check these and ask teachers to identify trivial segments - ask teachers to estimate task complexity of the segments for an inexperienced task-performer (with pre-specified prior knowledge) preferably by using an objective measurement instrument (following the procedure from Burtch et al.).	Identified functional segments
Step 1b –CTA phase Skill decomposition: Knowledge analysis	Identify for all segments the supportive knowledge; declarative knowledge that supports the performance of the non-recurrent aspects of the task. Again structured interviews are used to identify conceptual models, goal-plan models, causal models, and functional models. This is conducted parallel with other analysis. Only supportive knowledge identified for functional segments will be included. However during this analysis, functional segments are not yet known.The analysis results should be checked by reporting them to experts.	Identified supportive knowledge for functional segments

Continued

Appendix 1 (continued)

Step-description	*Activities (by instructional designer)*	*Outcomes*
Step 1c—CTA-phase Skill decomposition Scenario analysis	Identify more or less problem-dependent features in analyzing the various case studies gathered in the interviews. Report these features and cases to experts and ask teachers to order those cases on dimensions "complexity" and "representativeness".	Identification of possible problem situations for the instructional design phase.
Step 2—CTA-phase *Determination of task complexity*	Teachers use an instrument to measure the complexity for the tasks. Trivial tasks are excluded. Too simple tasks are also excluded for the further analysis.	Identification of the complexity for the functional tasks/segments.
Step 3—CTA-phase *SAP-analysis*	Identify from the interviews a juveline experts' (i.e., trainer) systematic approach to solve the task at hand and identify the heuristics in this approach. These can be several approaches and subset of them can be used in conjunction with functional tasks/segments. Report those SAPs and heuristics to experts and ask them about the "representativeness" of the SAPs. Finally ask teachers about the dimension "complexity" of identified SAPs.	Identified SAP and heuristics in which functional tasks/segments play a role.
Step 4—ID-phase *Micro-level sequencing*	Problem situations can be derived from case studies gathered in earlier steps. Look for variability along dimensions such as "the context in in which the task has to be performed", "the way of presenting the task", "the amount of support given when performing the task".	Identification of cases that adhere to variability of practice, working from examples to more general and abstract parts of knowledge and strategies.
Stap 5—ID-phase *Choosing problem formats*	The cognitive load theory and various other studies suggest to use "worked out examples" and problems with performance constraints combined with process worksheets (i.e., "process support problems" for instruction of ill-structured tasks such as in MmPs).	Problem formats for the problem situations in Step 4.
Step 6—ID-phase Determination of number of phases in process worksheets	Take prior knowledge and skills of learners into account, the complexity of the task assignments referred to in the SAP to be chosen should be comparable and neither too difficult nor too easy. For segments, this complexity is determined in earlier steps. Summarize the outcomes from Steps 4, 5 and 6 in a blueprint and ask experts to verify this overall outcome and agreement before the actual development of the instructional material takes place.	Number of phases for SAP to be presented in a process worksheet. Blueprint for MmP development combining outcomes from Steps 4, 5 and 6.

Note: ID= Instructional Design
SAP = systematic approaches to problem solving
MmP = Multimedia Practicals
CTA = cognitive task analysis

Appendix 2

Examples of to be rated tasks

very simple	simple	complex	very complex
Task formulation (to student) After celebrating in Amsterdam John is rather drunk and aggressive. As John is on his way home on December 12th 1995, someone gets in his way. He hits this person, Tim, hard on his nose. A while later, he runs into his old neighbour, who is also named Tim and this Tim also gets punched in the nose by John. Only the first Tim reports the assault to the police. John is charged as follows: "On or about December 12th in Amsterdam, a person named Tim was assaulted by John deliberately punching him in the nose. As a result he was painfully hurt. " Presented as evidence were a police report with the testimony of the (first) Tim and a testimony of a doctor, saying that after his examination on December 12th 1995, Tim was found to have a broken nose. John appears in court and states that he has no idea which fact the prosecutor is referring to.	*Task formulation (to student)* The Dutch trawler fleet consists of sixteen trawlers. They are fishing for herring, mackerel, and horsmackerel in the North Sea, but especially in more remote fishery grounds. This is regarded as 'big' seafishing, a healthy business which seldom attracts publicity. However, a short time ago a fight occurred on a Dutch traylor in the territorial waters of Denmark. In this fight, a member of the crew was badly injured by an oriental handweapon. The prosecuter procesutes the suspect on the charge of illegal possession of arms and attempted manslaughter, at least assault and battery. The suspect declares at the trial that he cannot be accused for this crime in the Netherlands, but he should stand trial in Denmark instead.	*Task formulation (to student)* During a football match between Ajax and Feyenoord, a Feyenoord player, Sjaak, is so irritated by the behavior of some Ajax-supporters, that he takes off his shoe and throws it at the supporters. A cleat on his shoe unfortunately hits Bram, one of the supporters, in the eye, damaging the retina. Sjaak is ejected and Bram needs treatment by his family doctor. He sends Bram to the hospital, to see an ophtalmologist. An operation is needed but there is a waiting list in the hospital which delays the operation for two months. Bram has been given instructions to visit the hospital on an empty stomach. In spite of this, Bram has eaten breakfast on the day of the operation causing a problem with the anesthesia. Bram dies during the operation.	*Task formulation (to student)* May 1999. The 42- year old Jacky Smith is paralysed from the neck down ever since her 68-year old mother shot her in March. Her mother, blind in one eye, was angry because she heard that her daughter was planning to move her to a home for the elderly. Because of this anger, she shot her daughter and attempted to shoot her daughter's boyfriend. The bullet hit the spiral cord causing the paralysis. She cannot swallow anymore and has to be fed artificially. According to her doctors, her situation is irreversible. She has received a judges permission to be allowed to die. Therefore, it is possible that her mother can be prosecuted for manslaughter. Suppose the situation above has occurred in Netherlands, and the victim uses her right to die by letting the medical apparatus be turned.
Question: What decision should be made by the judge? a. declare the subpoena invalid, because it does meet the criteria of art. 261 Sv. b. acquit John as the charge is unclear and cannot be proven. c. discharge John, as the fact cannot be qualified d. declare him guilty of charge if no exclusion of punishment presents itself	*Question:* Is the suspect correct or not? Motivate your answer.	*Question:* Motivate the chances of the prosecutor in in prosecuting Sjaak.	*Question:* Describe the chances of prosecuting the mother for manslaughter and also attend to arguments which could provide a conviction.

Art. 261 Sv refers to an article in Dutch Criminal Law. Answers are left out in these examples because of their substantive length. Original formulations were all in Dutch.

Summary
Workers struggle with rapid developments in their jobs which at the same time are becoming increasingly complex and where the solutions to the problems that they encounter are not readily available. Our society continues to make stronger demands on flexible problem solving behavior based upon the application of complex cognitive skills. Acquiring these skills can only be accomplished through complex learning that requires the integration of knowledge, skills, and attitudes as well as the coordination of qualitatively different constituent skills during task execution. Complex skills aim at transfer of what is learned to work settings or daily life. Typical examples are diagnosing a particular disease, selecting a suitable job applicant, modeling stress-factors that cause mental overload in workers, or preparing a plea to be held in court. The challenging question for education, thus, is: how can we help students acquire these complex cognitive skills?

Modern instructional theories focus increasingly on authentic learning tasks based on real-life tasks as the paramount condition for learning (e.g., Achtenhagen, 2001; Merrill, 2002; Reigeluth, 1999; van Merriënboer, 1997; van Merriënboer & Kirschner, 2001). A considerable risk with using authentic tasks is that they are often too difficult for novice learners to deal with as a whole. The most common solution, but one that is not optimal, is to split the ultimate skill into part-skills and teach them *separately* with the ultimate premise being that the learner will effectively combine them to achieve the desired end-state. A better solution to this problem is first to conceptually model reality (i.e., simplify it) and then to pedagogically model this model (Achtenhagen, 2001). Pedagogical modeling of the model (i.e., didactic specification, Resnick, 1976) is often achieved through the use of two *process support* mechanisms, namely (a) segmenting the whole learning task into smaller task assignments and thereby splitting the problem solving process into *phases* and presenting them, for example, via a process worksheet, and (b) offering *driving questions* to help learners in carrying out the activities within phases (Land, 2000). Process support is support for acquiring domain-based cognitive strategies.

Whole-task approaches (e.g., van Merriënboer, 1997) which focus on training - simple to complex versions of - the *complete* complex cognitive skill, emphasize the coordination and integration of the constituent skills from the very beginning and stress that learners should quickly develop a holistic vision of the whole task that is gradually embellished and detailed during the training.

Situated learning (Brown, Collins, & Duguid, 1989; Westera & Sloep, 1998) emphasizes that learning environments need to offer realistic situations where learning through meaningful practice takes place; the premise being that acquisition of complex skills is context-dependent and occurs most effectively in a relevant context (Anderson, 1982, 1993; Brown et al., 1989; Kirschner, van Vilsteren, Hummel, & Wigman, 1997; Kolb, 1984; van Parreren, 1987). Multimedia Practicals developed at the Open University of the Netherlands attempt to provide such realistic situations where meaningful practice takes place in an electronic *self-contained* learning environment (i.e., all necessary support is embedded in the environment).

Providing process support within a whole-task approach in Multimedia Practicals appears to be a fruitful way to foster the development of domain-based cognitive strategies. The amount of process support must, in turn, be optimized for efficient and effective learning. This requires the determination of objective task complexity, because complex tasks require more support than simpler tasks.

In this thesis a design approach for learning tasks within Multimedia Practicals is developed and tested, a task complexity measurement instrument for that approach is developed and validated, and two major questions on the provision of process support are studied, namely: (1) does the number of phases influence performance on the task and the efficiency of carrying out the task, and if so, in what way?, and (2) do driving questions positively influence performance on the task and the efficiency of carrying out the task?

Chapter 2 presents an Instructional Design model for Multimedia Practicals and its application. This model consists of two phases (cognitive task analysis and instructional design) both with three steps, and results in a whole-task instructional sequence with process support to facilitate the acquisition of a complex cognitive skill. Carrying out the design phase is based largely on the four-components Instructional Design-model (4C/ID-model; van Merriënboer, 1997) and insights from cognitive load theory (Sweller, 1988). One of the six steps requires the measurement of objective task complexity so as to determine the optimum number of phases.

Chapter 3 describes a developmental study that resulted in a reliable, valid and simple to use (i.e., not requiring specific training), four-category scale instrument for rating objective learning task complexity. This instrument resembles the Mohr-scale, a benchmark instrument for determining the hardness of minerals where a mineral is scratched with the 'anchor' minerals for comparison; the harder mineral leaves a scratch on the softer one. The benchmark instrument for determining task complexity, with anchor tasks representing the four categories, is used both in the Instructional Design model for Multimedia Practicals as well as for determining the complexity of the tasks in two empirical studies on process support. Advanced level graduate students were just as good in determining task complexity with the instrument as experienced teachers. The instrument delivers valid and reliable results and is particularly powerful for determining the extremes, but somewhat less powerful for determining the intermediate levels of complexity.

Chapter 4 describes the first study on process support which examined the effects of the number of phases (no phases, an intermediate number of phases, a high number of phases) on both task performance and task efficiency in order to shed light on the first research question: does the number of phases influence performance on the task and the efficiency of carrying out the task, and if so, in what way? Task efficiency is defined as task performance in relation to a combination of the mental effort, time on task, and level of motivation necessary to reach this level of performance. Sophomore Law students ($N = 35$) were randomly assigned to three versions of a Multimedia Practical that prepare them to carry out a plea in court. Students received whole-task training using three learning tasks, one of which was compulsory and two of which were additional, non-compulsory learning tasks. In addition, they had to execute one transfer task. It was hypothesized that students solving a whole learning task with an intermediate number of phases would outperform students exposed to either no phases at all (i.e., they learn to solve the whole task in only *one* phase) or a high number of phases. As hypothesized, an intermediate number of phases led to the most *effective* training showing that the number of phases could indeed be optimized for learning to carry out the whole learning task. As further expected, the condition with a high number of phases was least *efficient* for learning. But, contrary to the expectations, the intermediate number of phases condition was not most efficient for learning: no differences on time on task nor on task efficiency were found between the conditions with no-phases and an intermediate number of phases. As this study showed that the number of phases influences learning task performance and that a high number of phases is detrimental to task efficiency, logical steps were both to replicate the results of the number of phases and to study whether instructional design techniques *within* a phase, especially the provision of driving questions to help in carrying out the activities within phases, would influence task performance and task efficiency.

Chapter 5, reports on the second study on process support that examined the effects of the number of phases (low, high) and the availability of driving questions (available, not available) on task performance and task efficiency. The *low* number of phases in this study was comparable to the *intermediate* number of phases in the first study and the high number of phases was the same in both studies. As in the previous study, it was expected that students exposed to a lesser number of phases would outperform students exposed to a high number of phases and would be more efficient learners. Besides replicating the first study, the second study was also intended to provide an answer to the second major research question: do driving questions positively influence performance on the task and the efficiency of carrying out the task? Finally, this study also explored if an interaction effect occurs between the number of phases and the availability of driving questions on both task performance and task efficiency. Driving questions were expected to positively affect task performance and efficiency. For a possible interaction between the number of phases and driving questions it was expected that driving questions would have little added value if the number of phases is high (i.e., each phase is so simple that it can be easily carried out without help from driving questions) but become more valuable if the number of phases is low. In this study, sophomore Law students ($N = 82$) were randomly assigned to four versions of the aforementioned Multimedia Practical. As in the previous study, the whole task training uses three learning tasks (one of which was compulsory) as well as one transfer task. A main effect was found for the number of phases on the whole *learning task*, again showing that limiting the number of phases in learning to solve complex tasks leads to higher performance. A positive main effect for driving questions was found for plea inventory performance, one of the phases in the whole learning task. But the availability of driving questions, contrary to expectations, did not beneficially affect the efficiency of the students on the plea inventory. The absence of an effect of driving questions on both learning task performance and efficiency can be due to task characteristics (e.g., argumentative task domain, decreasing problem space for successive phases, convergent solution, interdependent phases) and/or feedback given at the completion of each phase. No interaction effects were observed, possibly because the difference between the low and high number of phases was too small to make the expected added value of driving questions manifest. The decreasing problem space and the provision of feedback after completion of each phase might also account for the absence of interaction effects.

Both process support studies showed, contrary to the expectations, that a high number of phases did not result in lowest performance and efficiency for the *transfer task*. This is probably due to the fact that students did not work on the non-compulsory learning tasks meant to offer variability of practice and so induce transfer of learning. The learning situation was thus limited to the compulsory learning task (i.e., one single, whole learning task). Researchers agree that transfer of learning cannot be expected under such circumstances (Gagné, Yekovich, & Yekovich, 1993; van Merriënboer, 1997).

Chapter 6 presents a general discussion in which the design approach and results are reviewed, followed by practical implications and guidelines, as well as suggestions for future research. It concludes with a final remark on ecological validity of the empirical studies.

In conclusion, although the application of the two-phase six-step Instructional Design model for Multimedia Practicals is resource-intensive in the cognitive task analysis phase, it should ultimately save costs during materials development and results in better maintainable, higher-quality materials. In addition to this, the results of the research on the task complexity instrument show that objectively determining task complexity prevents serious mistakes in the analysis and design phases, and also lowering costs as the pool of consultants for using this instrument increases because advanced level graduate students and experienced teachers rate similarly. In this, the Instructional Design model presented in this thesis clearly differs from other task-analytical methods that do not measure objective task complexity and almost always include the expensive and time-consuming use of 'true' experts.

The task complexity instrument makes it possible to optimize process support, and it allows developers to measure task complexity at reasonable costs prior to confronting learners with the learning environment so that the process support can be set up accordingly. Although the instrument is dedicated to measuring the complexity of Law learning tasks for sophomore Law students, the conceptual frame of reference used for its development offers a firm starting point for the development of analogous instruments for other domains.

The practical implications of the two empirical studies on process support are quite straightforward. There is clear empirical evidence for the value of offering process support through the distinction of a number of phases. Both studies showed that it is recommendable not to use too many phases. Too many phases leads to lower performance and does not justify the extra costs for developing such detailed instructional materials. Put bluntly, development costs can be reduced since less instructional material is needed. However, the ultimate reduction of the number of phases to a single-phase whole task, as seen in the first study, leads to lower performance. Thus, a further reduction in costs is precluded because a moderate number of phases should be distinguished to reach acceptable learning performance.

The finding that driving questions were probably only beneficial for early phases in problem solving, due to task characteristics (reduced problem space when proceeding in problem solving) and/or feedback given at the completion of each phase, implies that such driving questions might be useful, but that they can and should be faded during the learning process.

Finally, the findings can possibly be extended to domains with ontologies similar to that of Law.

In sum, these practical implications lead to the following design guidelines for process support in learning environments for the acquisition of complex cognitive skills:

a. determine the objective complexity of learning tasks
b. provide process support that takes task complexity into account
c. split the problem solving process of the whole task into phases, but not too many
d. consider providing driving questions in the early phases of problem solving
e. do all of this on the basis of a systematic design methodology, such as the two-phase six-step Instructional Design model for Multimedia Practicals.

Samenvatting
Beroepsbeoefenaren worstelen met de snelle ontwikkelingen binnen hun werk en de toenemende complexiteit daarvan, waarin ze in het werk tegen problemen aanlopen waarvoor geen pasklare oplossingen bestaan. Onze maatschappij doet een steeds groter beroep op flexibel probleemoplosgedrag dat is gebaseerd op de toepassing van complexe cognitieve vaardigheden. Het verwerven van deze vaardigheden kan alleen plaatsvinden door middel van complex leren. Complex leren vraagt om de integratie van kennis, vaardigheden en attitudes, en de coördinatie van kwalitatief verschillende samenstellende vaardigheden tijdens de uitvoering van een taak. Complexe vaardigheden zijn gericht op de transfer van het geleerde naar de werksituatie of het dagelijkse leven. Typerende voorbeelden daarvan zijn het diagnosticeren van een bepaalde ziekte, het selecteren van een geschikte sollicitant, het modelleren van stressfactoren die mentale overbelasting in het werk veroorzaken, of het voorbereiden van een pleidooi dat in de rechtszaal moet worden gehouden. De uitdagende onderwijsvraag is derhalve: hoe kunnen we studenten helpen bij het verwerven van dergelijke complexe cognitieve vaardigheden?

Recente instructietheorieën leggen steeds meer nadruk op het gebruik van authentieke leertaken gebaseerd op real-life taken als belangrijkste voorwaarde voor leren (e.g., Achtenhagen, 2001; Brown, Collins, & Duguid, 1989; Merrill, 2002; Reigeluth, 1999; van Merriënboer, 1997; van Merriënboer & Kirschner, 2001). Een aanzienlijk risico bij het gebruik van authentieke taken is dat deze voor beginners vaak te moeilijk zijn om als één geheel aan te pakken. De meest gebruikelijke, maar niet optimale oplossing is de gehele vaardigheid op te splitsen in deelvaardigheden en deze elk *afzonderlijk* te onderwijzen. Hierbij is de belangrijkste veronderstelling dat de lerende de deelvaardigheden uiteindelijk zelf effectief zal combineren om het gewenste eindniveau te bereiken. Een betere oplossing voor dit probleem is dat de werkelijkheid eerst conceptueel wordt gemodelleerd (i.e., wordt vereenvoudigd), en dat dit model vervolgens pedagogisch wordt gemodelleerd (Achtenhagen, 2001). Pedagogische modellering van het model (i.e., didactische specificatie, Resnick, 1976) wordt vaak bereikt door het gebruik van twee mechanismen voor *procesbegeleiding*, namelijk (a) het segmenteren van de hele leertaak in kleinere opdrachten waarbij het probleemoplosproces van de hele taak wordt opgesplitst in *fasen* die, bijvoorbeeld, in een proceswerkblad worden gepresenteerd, en (b) het verstrekken van *richtvragen* (Engels: driving questions) die de lerende helpen bij het uitvoeren van de activiteiten binnen fasen (Land, 2000). Procesbegeleiding is begeleiding gericht op het verwerven van domeingebonden cognitieve strategieën.

Hele-taak benaderingen (e.g., van Merriënboer, 1997) richten zich vooral op de training van eenvoudige tot complexe versies van de complexe cognitieve vaardigheid als *geheel*, benadrukken de coördinatie en integratie van de samenstellende vaardigheden vanaf het begin, en wijzen erop dat de lerende snel een holistische visie op de gehele taak dient te ontwikkelen die tijdens de training geleidelijk wordt aangekleed en gedetailleerd.

Leren in een context (Engels: situated learning) (Brown, Collins, & Duguid, 1989; Westera & Sloep, 1998) benadrukt dat leeromgevingen realistische situaties moeten bevatten waarbinnen leren via betekenisvolle oefening kan plaatsvinden. De veronderstelling is dat verwerving van complexe vaardigheden contextgebonden is en het meest effectief verloopt in een relevante context (Anderson, 1982, 1993; Brown et al., 1989; Kirschner, van Vilsteren, Hummel, & Wigman, 1997; Kolb, 1984; van Parreren, 1987). Multimediale Practica die bij de Open Universiteit Nederland zijn ontwikkeld proberen dergelijke realistische situaties te bieden waarbinnen betekenisvolle oefening kan plaatsvinden in een voor *zelfstudie* geschikte elektronische leeromgeving (i.e., alle vereiste begeleiding is in deze omgeving ingebouwd).

Het aanbieden van procesbegeleiding binnen een hele-taak benadering in Multimediale Practica lijkt een geschikte manier om het verwerven van domeingebonden cognitieve

strategieën te stimuleren. De hoeveelheid procesbegeleiding moet, op haar beurt, geoptimaliseerd worden met het oog op efficiënt en effectief leren. Dit vereist de bepaling van objectieve taakcomplexiteit omdat complexe taken om meer begeleiding vragen dan eenvoudige taken.
In dit proefschrift is een ontwerpbenadering voor leertaken in Multimediale Practica ontwikkeld en getoetst, is een instrument voor het meten van taakcomplexiteit binnen deze benadering ontwikkeld en gevalideerd, en zijn twee hoofdvragen bestudeerd bij het aanbieden van procesbegeleiding, namelijk: (1) beïnvloedt het aantal fasen de prestatie op de taak en de efficiëntie van het uitvoeren van de taak, en zo ja, op welke manier?, en (2) hebben richtvragen een positieve invloed op de prestatie op de taak en op de efficiëntie van het uitvoeren van de taak?

Hoofdstuk 2 presenteert een instructie-ontwerpmodel voor Multimediale Practica en diens toepassing. Dit model bestaat uit twee fasen (cognitieve taakanalyse en instructie-ontwerp) met elk drie stappen. Toepassing leidt tot een sequentie van hele leertaken met procesbegeleiding die het verwerven van een complexe cognitieve vaardigheid ondersteunt. Het uitvoeren van de ontwerpfase is voor een belangrijk deel gebaseerd op het vier-componenten instructie-ontwerpmodel (4C/ID-model; van Merriënboer, 1997) en op inzichten uit de cognitieve belastingstheorie (Sweller, 1988). Een van de zes stappen vraagt om de meting van objectieve taakcomplexiteit om het optimale aantal fasen te kunnen bepalen.

Hoofdstuk 3 beschrijft een ontwikkelingsonderzoek dat heeft geleid tot een betrouwbaar, valide en eenvoudig te gebruiken (i.e., geen specifieke training nodig) instrument voor het bepalen van de objectieve complexiteit van leertaken. De schaal van dit instrument telt vier categorieën. Dit instrument lijkt op de Mohr-schaal, een ijkinstrument (Engels: benchmark instrument) voor het bepalen van de hardheid van gesteenten, waarbij het gesteente wordt bekrast door zogenaamde 'anker' gesteenten; het hardere gesteente laat een kras achter op het zachtere gesteente. Het ijkinstrument voor het bepalen van taakcomplexiteit heeft ankertaken die representatief zijn voor de vier categorieën. Dit ijkinstrument is gebruikt bij het instructie-ontwerpmodel voor Multimediale Practica en de bepaling van de complexiteit van de taken in twee empirische studies naar procesbegeleiding. Bijna afgestudeerde studenten waren in staat om net zo goed als docenten taakcomplexiteit te bepalen met dit instrument. Het instrument levert valide en betrouwbare resultaten en blijkt vooral krachtig om de extremen te bepalen, maar minder krachtig bij het bepalen van de tussengelegen niveaus van complexiteit.

Hoofdstuk 4 beschrijft de eerste studie naar procesbegeleiding waarin het effect van het aantal fasen (geen fase, een tussengelegen aantal fasen, een hoog aantal fasen) op zowel taakprestatie als op taakefficiëntie is nagegaan. Dit diende licht te werpen op de eerste onderzoeksvraag: beïnvloedt het aantal fasen de prestatie op de taak en de efficiëntie van het uitvoeren van de taak, en zo ja, op welke manier? Taakefficiëntie is gedefinieerd als taakprestatie in relatie tot een combinatie van de benodigde mentale inspanning, studielast en niveau van motivatie om dit niveau van taakprestatie te bereiken. Tweedejaars studenten Rechten ($N = 35$) zijn aselect toegewezen aan drie versies van het Multimediale Practicum dat hen voorbereid op het houden van een pleidooi in de rechtszaal. De studenten kregen een hele-taak training die uit drie leertaken bestond, waarvan één verplicht en waarvan twee facultatieve leertaken. Daarnaast moesten ze een transfer taak uitvoeren. De veronderstelling was dat studenten die een hele taak met een tussengelegen aantal fasen zouden oplossen beter zouden presteren dan studenten die werden geconfronteerd met geen fasen (i.e., zij leerden om de hele taak in *één* fase op te lossen) danwel een hoog aantal fasen. Zoals verondersteld leidde het tussengelegen aantal fasen tot de meest *effectieve* training, hiermee aantonend dat het aantal fasen voor het leren oplossen van de hele taak inderdaad kon worden geoptimaliseerd. Zoals verder verwacht bleek de conditie met een hoog aantal fasen het minst *efficiënt* voor het leren. Maar in tegenstelling tot de verwachtingen bleek het tussengelegen aantal fasen niet het meest efficiënt voor het leren: de condities zonder fasen en met een tussengelegen aantal fasen verschillen niet in studielast en taakefficiëntie. Deze studie toonde aan dat het aantal fasen de prestatie op de leertaak beïnvloedt en een hoog aantal fasen een nadelige invloed heeft op taakefficiëntie. Derhalve waren het logische vervolgstappen om zowel deze resultaten van het aantal fasen te repliceren, alsmede te onderzoeken of instructiemaatregelen *binnen* een fase, in het bijzonder het aanbieden van richtvragen om te helpen bij het uitvoeren van de activiteiten binnen een fase, invloed zouden hebben op taakprestatie en taakefficiëntie.

Hoofdstuk 5 rapporteert over de tweede studie naar procesbegeleiding waarin het effect van het aantal fasen (laag, hoog) en de beschikbaarheid van richtvragen (beschikbaar, niet beschikbaar) op taakprestatie en taakefficiëntie werd nagegaan. Het *lage* aantal fasen in deze studie was vergelijkbaar met het *tussengelegen* aantal fasen in de eerste studie, terwijl het hoge aantal fasen in beide studies identiek was. Zoals in de vorige studie, er werd verwacht dat de studenten die met een kleiner aantal fasen werden geconfronteerd beter zouden presteren en bovendien efficiënter zouden leren dan de studenten die aan een hoog aantal fasen werden blootgesteld. Behalve als replicatie van de eerste studie was de tweede studie ook bedoeld om een antwoord te geven op de tweede onderzoeksvraag: hebben richtvragen een positieve invloed op de prestatie op de taak en op de efficiëntie van het uitvoeren van de taak? Ten slotte exploreerde deze studie of er een interactie-effect voorkomt tussen het aantal fasen en de beschikbaarheid van richtvragen op zowel taakprestatie alsook taakefficiëntie. Er werd verwacht dat richtvragen een positieve invloed zouden hebben op taakprestatie en taakefficiëntie. Voor een mogelijk interactie-effect tussen aantal fasen en richtvragen was de verwachting dat richtvragen weinig toegevoegde waarde zouden hebben indien het aantal fasen hoog is (i.e., elke fase is zo eenvoudig dat deze eenvoudig zonder de hulp van richtvragen kan worden uitgevoerd) maar belangrijker zouden worden indien het aantal fasen lager is. In deze studie zijn tweedejaars studenten Rechten ($N = 82$) aselect toegewezen aan vier versies van het eerder genoemde Multimediale Practicum. Net als in de vorige studie bestaat de hele-taak training uit drie leertaken (waarvan één verplicht) alsmede een transfer taak. Een hoofdeffect werd gevonden voor het aantal fasen op de hele *leertaak*, opnieuw aantonend dat het beperken van het aantal fasen bij het leren oplossen van complexe taken tot een betere prestatie leidt. Er werd een positief hoofdeffect van het aanbieden van richtvragen gevonden op de prestatie op de opdracht 'vertrouwd raken met het dossier' (i.e., pleitoverzicht), een van de fasen in de hele leertaak. De beschikbaarheid van richtvragen had echter, in tegenstelling tot de verwachtingen, geen positieve invloed op de efficiëntie van de studenten op het pleitoverzicht. De afwezigheid van een effect van richtvragen op zowel de prestatie op de leertaak alsook op de efficiëntie kan het gevolg zijn van de taakkarakteristieken (e.g., argumentatief taakdomein, kleiner wordende probleemruimte bij opeenvolgende fasen, convergente oplossing, onderling afhankelijke fasen) en/of van de feedback die bij het afronden van elke fase wordt gegeven. Er werden geen interactie-effecten geconstateerd, waarschijnlijk omdat het verschil tussen het lage en hoge aantal fasen te klein was om de verwachte meerwaarde van richtvragen zichtbaar te maken. De afwezigheid van interactie-effecten zou ook verklaard kunnen worden door de kleiner wordende probleemruimte en de verstrekking van feedback bij het afronden van elke fase.

Beide studies naar procesbegeleiding lieten, in tegenstelling tot de verwachtingen, zien dat een hoog aantal fasen niet leidt tot de laagste prestatie en efficiëntie op de *transfer-taak*. Waarschijnlijk is dit veroorzaakt door het feit dat de studenten niet met de facultatieve leertaken hebben gewerkt. Deze taken waren bedoeld om een variatie in oefening te bieden en zodoende transfer te stimuleren. De leersituatie was dus beperkt tot de verplichte leertaak (i.e., één enkele hele leertaak). Onderzoekers zijn het eens dat transfer van leren in dergelijke omstandigheden niet kan worden verwacht (Gagné, Yekovich, & Yekovich, 1993; van Merriënboer, 1997).

Hoofdstuk 6 presenteert de algemene discussie waarin de ontwerpbenadering en de resultaten worden besproken, gevolgd door praktische aanbevelingen en richtlijnen, en suggesties voor verder onderzoek. Het eindigt met een slotopmerking over ecologische validiteit van de empirische studies.

Concluderend kan worden gesteld dat hoewel de toepassing van het twee-fasen zes-stappen instructie-ontwerpmodel voor Multimediale Practica arbeidsintensief is in de fase van de cognitieve taakanalyse, het bij de ontwikkeling van leermaterialen uiteindelijk kostenbesparend zou moeten zijn en kan leiden tot beter onderhoudbare en kwalitatieve leermaterialen. Daarnaast toonden de uitkomsten van het onderzoek naar het instrument voor het bepalen taakcomplexiteit aan dat objectieve bepaling van taakcomplexiteit ernstige fouten in de analyse - en ontwerpfasen helpt voorkomen, en ook kostenbesparing oplevert omdat het aantal deskundigen dat dit instrument kan gebruiken toeneemt aangezien bijna afgestudeerde studenten en ervaren docenten op vergelijkbare wijze blijken te schatten. Hierdoor wijkt het in dit proefschrift gepresenteerde instructie-ontwerpmodel duidelijk af van andere taakanalytische methodes, die niet objectieve taakcomplexiteit bepalen en vrijwel altijd kostbaar en tijdsintensief gebruik maken van 'echte' deskundigen.

Een instrument voor het meten van objectieve taakcomplexiteit maakt optimalisatie van procesbegeleiding haalbaar, en maakt het voor ontwikkelaars mogelijk tegen redelijke kosten taakcomplexiteit te meten en procesbegeleiding dienovereenkomstig op te zetten voordat lerenden met de leeromgeving in contact worden gebracht. Het ontwikkelde instrument is geschikt voor het meten van de complexiteit van leertaken Recht die aan tweedejaars studenten Rechten worden voorgelegd, maar het conceptuele kader voor de ontwikkeling van dit instrument vormt een goed vertrekpunt voor het ontwikkelen van analoge instrumenten voor andere domeinen.

De praktische aanbevelingen van de twee empirische studies naar procesbegeleiding zijn vrij duidelijk. Er is duidelijke empirisch steun voor de waarde van het verstrekken van procesbegeleiding door een onderscheid in het aantal fasen. Beide studies hebben aangetoond dat het raadzaam is niet teveel fasen te gebruiken. Het gebruik van teveel fasen leidt tot een slechtere prestatie en rechtvaardigt niet de extra kosten voor het ontwikkelen van dergelijke gedetailleerde leermaterialen. Onomwonden gesteld, ontwikkelkosten kunnen worden gereduceerd omdat minder leermateriaal nodig is. De ultieme reductie van het aantal fasen tot een hele taak die uit één enkele fase bestaat, zoals bij de eerste studie het geval was, leidt echter tot een lagere prestatie. Een verdere beperking in de kosten is uitgesloten omdat toch een zeker aantal fasen nodig blijkt te zijn om een acceptabele leerprestatie te bereiken.

De constatering dat richtvragen mogelijkerwijs alleen nuttig zijn voor fasen in het begin van het probleemoplossen, is te wijten aan taakkarakteristieken (een kleiner wordende probleemruimte bij het verder gaan met het oplossen van het probleem) en/of terugkoppeling bij het voltooien van elke fase, en impliceert dat richtvragen zinvol kunnen zijn, maar dat ze tijdens het leerproces kunnen en moeten afnemen.

Ten slotte kunnen de bevindingen waarschijnlijk worden uitgebreid tot domeinen met ontologieën die vergelijkbaar zijn met dat van Rechten.

Samenvattend leiden deze praktische aanbevelingen tot de volgende ontwerprichtlijnen voor procesbegeleiding in leeromgevingen bij het verwerven van complexe cognitieve vaardigheden:

a. bepaal de objectieve complexiteit van leertaken
b. geef procesbegeleiding die rekening houdt met de taakcomplexiteit
c. splits het probleemoplosproces van een hele taak in fasen, maar niet teveel
d. overweeg richtvragen te verstrekken in de fasen aan het begin van het probleemoplosproces
e. doe al het bovenstaande op basis van een systematische ontwerpmethode zoals het twee-fasen zes-stappen instructie-ontwerpmodel voor Multimediale Practica.

Dankwoord

Bij dit proefschrift hebben veel mensen hun support geleverd die ik hier graag wil bedanken.

Allereerst natuurlijk mijn promotoren Jeroen van Merriënboer en Paul Kirschner, tevens mijn "dagelijks" begeleider. Jullie hebben mij via double-faded process support geleidelijk in de wereld van de wetenschap ingewijd. Het rijden van deze "Tour de Science" hebben jullie netjes in etappes opgesplitst en bovendien driving questions verstrekt voor de afzonderlijk te rijden etappes. Alsof dat nog niet genoeg was hebben jullie ook de product support geleverd die in onze gezamenlijke artikelen is neergeslagen. Dankzij jullie heb ik mijn plek in het peloton der onderzoekers gevonden.

Daarnaast heb ik van velen product support en enabling support ontvangen. Deelnemers, ravitaillering, materiaal, massages en wat dies meer zij; het is allemaal nodig om etappes van start te laten gaan en het uitrijden van de Tour mogelijk te maken.

(1) Vooral dankzij Aad Slootmaker, Jürgen Wöretshofer en Henk van den Brink is al het studiemateriaal voor het onderzoek tot stand gekomen. Het studiemateriaal is gebaseerd op *Pleit voorbereid* waarbij naast de reeds genoemde personen ook Tonnie Starren, Lieke Quanjel, Carlo Aretz, Natasja van der Meer en George Martijn een essentiëel aandeel in de ontwikkeling van de inhoud hebben gehad en waarbij Brigitte De Craene mede aan de onderwijskundige basis heeft gestaan. Studenten gebruikten dit materiaal voor het verwerven van de pleitvaardigheid.

(2) Dankzij Martin Baks, Dick van Ekelenburg, en Jürgen Wöretshofer zijn de leertaken ontwikkeld en geselecteerd waarvan docenten Rechten en bijna afgestudeerde studenten Rechten hebben aangegeven hoe complex deze zouden zijn voor tweedejaars studenten Rechten om ze uit te voeren.

(3) Docenten Rechten en bijna afgestudeerde studenten Rechten hebben bijgedragen aan de ontwikkeling van het instrument voor het objectief meten van complexiteit van leertaken. Hun beoordelingen van complexiteit zijn ons een enorme steun geweest bij de ontwikkeling van dit meetinstrument en van het studiemateriaal voor het onderzoek.

(4) Dankzij Lisette Boeren, Tonnie Starrren, Sylvia Walther, Martin Baks en Hans Hummel zijn de instrumenten ontwikkeld waarmee de producten zijn beoordeeld die de studenten bij de verwerving van hun pleitvaardigheid aanmaakten.

(5) Dankzij Loet van Wijk, Anja van Valen, José Plug, Robin de Roon, Harm Kloosterhuis, Ankie Broekers-Knol, Helmi de Ruiter, Ingrid van den Oord, Matthea Verdaasdonk, Natasja van der Meer, Carlo Aretz en Anita Kessen zijn de studenten gevonden die vaardige pleiters wilden worden en dankzij Renée Lemaire is dit ook juridisch geregeld.

(6) Dankzij Cisca Andeweg, Ingrid Jonkman, Nicole Knebel, Brigitte Peters en Alex Ruis is al het in het onderzoek gebruikte materiaal in goede staat bij de deelnemers verschenen.

(7) Dankzij de studenten die zich zonder te protesteren tot de pleitvaardigheid wilden bekwamen hebben we de spirit gekregen om de etappes en de Tour uit te rijden.

(8) Dankzij Aad Slootmaker zijn alle gegevens uit het door studenten elektronisch geretourneerde studiemateriaal gedestilleerd.

(9) Vooral door Cornelie Arnouts, Susanne Munsters, Johan Tisscher en Justus Faber zijn de producten beoordeeld die de hierboven genoemde studenten aanmaakten. Ook Fleur Landa en vrijwel alle bij (5) en (4) genoemde personen hebben hierin een aandeel gehad.

(10) Mede dankzij Nick Broers en Hans van Buuren konden de gegevens statistisch worden geanalyseerd.

(11) Dankzij Bob Wilkinson en Paul Kirschner is de verslaglegging in het Engels verbeterd.

(12) Dankzij Jeroen Storm en Jeroen Berkhout zijn de afbeeldingen op en in het proefschrift tot stand gekomen.

Alle bij (1) tot en met (12) genoemde personen wil ik tevens heel hartelijk danken voor de prettige manier waarop ze medewerking hebben verleend.

Fellow support heb ik vooral van mijn paranimfen Hans Hummel en Aad Slootmaker ontvangen. Gezamenlijk hebben we de toppen en dalen in het parcours beleefd en elkaar daarin ook inhoudelijk uitgedaagd. Hans en Aad, another Tour to go!

Financial support is van essentieel belang, zonder sponsor geen Tour. Ik wil in het bijzonder Wim Jochems en Freek Gastkemper bedanken voor het mogelijk maken van dit onderzoek en hun in mij gestelde vertrouwen. Tevens mijn dank aan het managementteam van de faculteit Rechtswetenschappen van de Open Universiteit Nederland voor het laten participeren van docenten Rechten in dit onderzoek. Het geld dat jullie hiertoe van de sponsor ontvingen heeft ongetwijfeld een nobel en rechtvaardig doel gevonden.

Virtual support, dat wil zeggen voor lief nemen dat ik voor 40 % van mijn werktijd voor promotieonderzoek was "ondergedoken", is door de collega's in diverse interne en externe implementatieprojecten en door overige OTEC-collega's geleverd. Ik ben jullie daar zeer dankbaar voor. Zo nam ik ook de "rust" om de etappes te rijden.

Emotional support heb ik van (OTEC)-collega's, vrienden en familie gekregen. Om geen van hen te kort te doen noem ik liever geen namen apart. Maar emotional support heb ik vooral gekregen van mijn geliefde en privé-shrink Monique Wijers. Zonder haar was ik wellicht toch in de bezemwagen gestapt.

Ten slotte, unconditional support viel mij ten deel als onze zoon Sam op zijn eigen manier duidelijk maakte waar het feitelijk om draait in het leven.
Ik draai weer mee!

Maastricht, januari 2004.

Curriculum Vitae

Rob Nadolski (1961) is na afronding van een atheneum B opleiding en de lerarenopleiding in de vakken wiskunde en natuurkunde vanaf 1984 betrokken bij projecten waarin de ontwikkeling van computer ondersteund onderwijs (COO) centraal staat. Tot 1988 bij het Centrum voor Onderwijs en Informatietechnologie en het Educational Computing Consortium in Enschede en daarna tot heden bij de Open Universiteit Nederland te Heerlen alwaar hij vooral het ontwerp en de projectleiding binnen multidisciplinair samengestelde teams bij dergelijke projecten doet.